UNITÉ DE RECHERCHE
INRIA-ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P. 105
78153 Le Chesnay Cedex
France
Tél.:(1) 39 63 55 11

Rapports de Recherche

N° 1508

Programme 5
Traitement du Signal,
Automatique et Productique

NEW VARIANTS OF BUNDLE METHODS

Claude LEMARÉCHAL
Arkadii NEMIROVSKII
Yurii NESTEROV

Septembre 1991

RR.1508

New Variants of Bundle Methods

Nouvelles variantes des methodes de faisceaux

Claude Lemaréchal*, Arkadii Nemirovskii**, Yurii Nesterov**

In this paper, we study bundle-type methods for convex optimization, based on successive approximations of the optimal value. They enjoy optimal efficiency estimates; furthermore, they provide attractive alternatives to solving convex constrained optimization problems, convex-concave saddle-point problems, and variational inequalities. We present a number of possible variants, establish their efficiency estimate, and give some illustrative numerical results.

Cet article concerne des méthodes de type faisceaux pour l'optimisation convexe, construisant des approximations successives de la valeur optimale. Leur vitesse de convergence est optimale; de plus elles fournissent d'intéressantes méthodes pour le cas contraint, les problèmes de point-selles, et les inégalités variationnelles. Nous présentons plusieurs variantes possibles, établissant leur vitesse de convergence, et nous les illustrons sur quelques exemples numériques.

Mots-clés: optimisation non différentiable, points-selles, inégalités variationnelles, méthodes de faisceaux.

Classification AMS: 49-02, 49D37, 65-02, 65K05, 90-02, 90C30.

* INRIA, BP 105, 78153 Le Chesnay, France (lemarech@seti.inria.fr).

** CEMI, 32 Krasikova, Moscou 117481, URSS (root@cemi.msk.su).

CONTENTS

0. Introduction	1
1. Problems	9
2. Methods for (Min)	11
3, Methods for (Sad)	23
4. Methods for (CMin)	35
5. A Method for (Var)	48
6. Computational results	51
References	57
Appendix 1	60
Appendix 2	62

0. Introduction

0.1. Consider the basic problem of minimizing a convex function f over a "simple" convex set $Q \subset \mathbb{R}^n$. Having generated the iterates $x_1,\dots,x_i \in Q$ and using an *oracle* to compute function-values $f(x)$ and subgradient-values $f'(x)$, a fruitful object is the *cutting-plane model*

$$f_i(x) = \max\{f(x_j) + (f'(x_j))^T(x-x_j) \mid 1 \le j \le i\}$$

under-estimating f. To exploit it, the very first idea is the classical *cutting-plane algorithm* of [Ke. 1960], [CG 1959], in which x_{i+1} minimizes f_i over Q; it is known as very slow, both from the theoretical and practical viewpoints; see [NYu 1983] for example.

More recently, some refinements of this idea have been proposed, under the wording of *bundle methods*. In their simplest form [Le. 1978], [Mi. 1982], [Ki. 1983], the next iterate is

$$x_{i+1} = \operatorname{argmin}\{f_i(x) + \frac{1}{2} u_i \, |x-x_i^+|^2 \mid x \in Q\} \qquad (0.1)$$

where the *current prox-center* x_i^+ is a certain point from the set $\{x_1,\dots,x_i\}$ and u_i is the *current penalty parameter*. If $f(x_{i+1})$ turns out to have "sufficently decreased" (*descent step*), the prox-center is updated to x_{i+1}; otherwise (*null step*), $x_{i+1}^+ = x_i^+$. This idea looks natural: the model accumulates all the information about f obtained so far, and the penalty term reduces the influence of the model's inaccuracy, thereby reducing instabilities. A bundle method is thus determined by two rules: (1) to define a "sufficient" decrease, and (2) to select the penalty parameter. Satisfactory rules have been developed for (1), based on a comparison between the actual value $f(x_{i+1})$ and the "ideal" value

$f_i(x_{i+1})$ of the model. As for (2), the question is not so clear: the simplest choice $u_i \equiv 1$ is theoretically possible but experience demonstrates that efficiency requires "on line" adjustments, as in [Ki. 1990], [SZ 1991].

0.2. Alternatives to (0.1) can be considered, which have the same stabilizing effect. Let us mention two of them: the "trust-region approach"

$$x_{i+1} = \text{argmin}\{f_i(x) \mid x \in Q,\ |x-x_i^+| \leq \tau_i\},$$

which does not seem to have been studied, and the proposal of [LSB 1981], in which the control parameter is a certain ε_i, whose choice implies a detour in the dual space. In what follows, we study a fourth variant: instead of u_i, τ_i or ε_i, we control the value of the model at the next iterate: we choose a *level* l_i and replace (0.1) by

$$x_{i+1} = \text{argmin}\{\tfrac{1}{2}|x-x_i^+|^2 \mid x \in Q,\ f_i(x) \leq l_i\}. \tag{0.2}$$

It turns out that the level-sets of the model are rather "stable", so that extremely simple rules can be used for updating the level l_i. This property also allows us to forget about the concepts of prox-center and null-step: x_i^+ may be systematically set to the last iterate x_i in (0.2).

Our basic strategy works as follows: at the i-th step, compute the minimal value $f_*(i)$ of the model over Q (assumed bounded); also, let

$$f^*(i) = \min\{f(x_j) \mid 1 \leq j \leq i\} = f(x_i^*)$$

be the best value of the objective obtained during the first i steps, and call

$$\Delta(i) = f^*(i) - f_*(i) \tag{0.3}$$

the i-th *gap* (x_i^* certainly minimizes f within $\Delta(i)$ and our aim is to force $\Delta(i) \to 0$). Then, having $\lambda \in (0,1)$, solve (0.2) with the value

$$l_i = \lambda f^*(i) + (1-\lambda) f_*(i) = f_*(i) + \lambda \Delta(i). \tag{0.4}$$

0.3. Needless to say, the value $\lambda = 1$ in (0.4) would result in $x_{i+1} = x_{i+1}^+$; a value close to 1 would mimic a pure subgradient method with very short steps, possibly converging to a wrong point. By contrast, $\lambda = 1$ would yield the convergent (even though slow) pure cutting plane methods; this suggests that small values should be less dangerous than large values of λ, i.e., of the level.

An arbitrary but fixed $\lambda \in (0,1)$ gives the following efficiency estimate: to obtain a gap smaller than ε, it suffices to perform

$$M(\varepsilon) \le c\ (LD/\varepsilon)^2 \tag{0.5}$$

iterations (here, L and D are the Lipschitz constant of f and the diameter of Q respectively, c is a constant depending only on λ). Such an efficiency is optimal in a certain sense (see [NYu 1983]): suppose Q is a ball of radius $D/2$, the dimension is $n \ge 4^{-1}(LD/\varepsilon)^2$, take an arbitrary method but use at most $4^{-1}(LD/\varepsilon)^2$ evaluations of f and f' (and no other information from the problem); then, there exists a function for which this method does not obtain an accuracy better than ε. As a result, our method cannot be improved uniformly with respect to the dimension by more than an absolute constant factor.

To obtain the estimate (0.5), the key argument is as follows: consider, for a given i_0, the maximal sequence $I =$

$\{i_0, i_0+1, \ldots, i_1\}$ of iterations (we call it a *group*), at the end of which the gap has not been reduced much, namely,

$$\Delta(i_1) \geq (1-\lambda)\, \Delta(i) \text{ for all } i \in I.$$

Then, all level-sets characterising (0.2) with $i \in I$ have a point in common. This crucial property allows the following majoration of the number of iterations in the group:

$$|I| \leq c(LD/\Delta(i_1))^2,$$

where c is a constant depending only on λ. Then, using the fact that the gap is reduced by $(1-\lambda)$ at the iteration i_1+1, repeated use of this argument results in the majoration (0.5).

In Section 2 we present a number of variants of the above algorithm, all enjoying the same efficiency property (0.5).

0.4. The subsequent sections are devoted to problems for which the same idea can be considered. After all, the above "level" principle gives an implementable mechanism to solve a system of inequations (via a method resembling Newton's method, see [Ro. 1972]): we want to find $x \in Q$ such that

$$f(x') + (f'(x'))^T(x-x') \;[\leq f(x)] \leq f^* \quad \text{for all} \quad x' \in Q.$$

Here, there are infinitely many indices, so they are accumulated one after the other: $x' = x_1, x_2, \ldots$; and f^* is unknown, so the level-strategy takes care of it.

The essential feature to make the method work is to define an appropriate nonnegative *gap* as in (0.3), which is 0 when the problem is solved. The whole approach is therefore to minimize this gap, an idea which can actually be extended to several problems.

A. Saddle-point problems (Section 3): given a convex-concave function $f(x,y)$ defined on the direct product of Q and H (convex and compact), find a *saddle point* $(x^*,y^*) \in Q{\times}H$, i.e. a point satisfying

$$\max\{f(x^*,y) \mid y \in H\} = f(x^*,y^*) = \min\{f(x,y^*) \mid x \in Q\}.$$

This just amounts to minimizing the convex function

$$F(x,y) = \max_H f(x,\cdot) - \min_Q f(\cdot,y)$$

over $Q{\times}H$. The difficulty is that we have no oracle computing the values and the subgradients of F; nevertheless, a set of iterates $\{(x_j,y_j) \mid 1 \le j \le i\}$ yields the *model*

$$F_i(x,y) = \overline{f}_i(x) - \underline{f}_i(y), \tag{0.6}$$

where the standard first-order information is used:

$$\overline{f}_i(x) = \max\{f(x_j,y_j) + (f'_x(x_j,y_j))^T(x-x_j) \mid 1 \le j \le i\},$$

$$\underline{f}_i(y) = \min\{f(x_j,y_j) + (f'_y(x_j,y_j))^T(y-y_j) \mid 1 \le j \le i\};$$

thus, F_i underestimates F. We know that the minimal value of F is zero; the minimal value of each F_i is therefore nonpositive and provides the gap $\Delta_i=-F_i$. This enables us to define a method of the type (0.2) for saddle-point problems with the efficiency estimate (0.5).

It is interesting to note the decomposed property of the model (0.6): to minimize it, it suffices to solve successively the two linearized optimization problems

$$\min_Q \overline{f}_i(x) \quad \text{and then} \quad \max_H \underline{f}_i(y).$$

This suggests an interpretation of our approach in terms of games: there are two players x and y, in charge of minimizing f and $-f$, respectively; $\overline{f}_i$ and $(-\underline{f}_i)$ can be interpreted as under-approximations of their worst-case loss-functions.

We recall that the usual algorithms for saddle-points are based on subgradient optimization [Er. 1966]. In [Au. 1972], approaches similar to ours were considered, but of course based on pure cutting-plane approximations.

B. Convex constrained problems (Section 4). Given the function G, convex on the compact convex set Q, our approach to solve

$$\min\{f(x) \mid G(x) \leq 0,\ x \in Q\}$$

is via the equivalent problem

$$\min\{\max[f(x)-f^*, G(x)] \mid x \in Q\}. \tag{0.7}$$

(it is to alleviate notations that we assume just one inequality constraint). The optimal value f^* is of course unknown, which introduces a new difficulty: no oracle can compute the function-value in (0.7). We therefore under-estimate f^* by the optimal value $f_*(i)$ of

$$\min\{f_i(x) \mid G_i(x) \leq 0,\ x \in Q\}$$

(G_i being the cutting-plane approximation of G), and we propose two approaches.

First, duality theory tells us that (0.7) is equivalent to

$$\max\{h(\alpha)-\alpha f^* \mid 0 \leq \alpha \leq 1\} \tag{0.8}$$

where

$$h(\alpha) = \min\{\alpha f(x) + (1-\alpha)G(x) \mid x \in Q\}$$

can be over-estimated by the function

$$h_i(\alpha) = \min\{\alpha f(x_j) + (1-\alpha)G(x_j) \mid 1 \leq j \leq i\}.$$

Thus, a gap is obtained:

$$\Delta_i = \max\{h_i(\alpha) - \alpha f^*(i) \mid 0 \leq \alpha \leq 1\}$$

which must be reduced to the optimal value in (0.8), i.e. in (0.7), namely 0.

In our second approach, f^* is replaced by a parameter t, and the problem is to solve the equation

$$\kappa(t) = \min\{\max\{f(x)-t, G(x)\} \mid x \in Q\} = 0$$

(this is close to the method of "loaded functional" [Lb. 1977]). Here again, κ cannot be computed exactly. A gap is therefore defined, by way of cutting-plane approximations in κ, and t is updated to the current $f_*(i)$ whenever this gap diminishes by a sufficient amount.

In both methods, the need to identify f^* while solving the saddle point problem (0.7) is paid by an extra cost in the efficiency estimate, which becomes as follows: to reach a point x satisfying

$$f(x) \le f^* + \varepsilon \qquad \text{and} \qquad G(x) \le \varepsilon,$$

its suffices to perform

$$M(\varepsilon) \le c\ (LD/\varepsilon)^2 \ln(LD/\varepsilon)$$

iterations. Note, however, that no Slater assumption is needed; as a result, the efficiency is not affected by large Lagrange multipliers, as is the case with methods involving exact penalty.

C. Variational inequalities with monotone operators (Section 5) also admit a solution procedure of the type (0.2) with efficiency estimate (0.5). Indeed, consider again Section 0.1: in the definition of the model f_i, replace the values $f(x_i)$ by the current best value $f^*(i)$. The result is a further underestimate of the model:

$$\phi_i(x) = f^*(i) + \max\{(f'(x_i))^T(x-x_i) \mid 1 \le j \le i\} \le f_i(x),$$

so a variant of the level algorithm is readily obtained if we replace the function f_i by ϕ_i (note the similarity with the *conjugate subgradient* approach of [Le. 1975], [Wl. 1975]). The interest of this variant is that function-values are no longer involved, so it can be used to solve the problem

$$\text{find } x \in Q \text{ s.t. } \langle F(x'),x'-x\rangle \geq 0 \quad \text{for all } x' \in Q \tag{0.9}$$

(F is a (possibly multivalued) monotone mapping and Q is again closed and convex). Here, the monotone mapping F plays the role of f' and ϕ_i allows the definition of a gap Δ_i associated with the function

$$f(x) = \sup\{\langle F(x'),x-x'\rangle \mid x' \in Q\}.$$

The resulting method is reminiscent of [MD 1989], but continuity of $F(\cdot)$ is not assumed (although we require both F and Q to be bounded).

Recall that the standard formulation of a variational inequality is

$$\text{find } x \in Q \text{ s.t. } \langle F(x),x'-x\rangle \geq 0 \text{ for all } x' \in Q, \tag{0.10}$$

which *is not* the same as (0.9). It can be proved, however, that (0.9) and (0.10) are equivalent in the maximal monotone case (see Appendix for precise formulations).

An important computational advantage of (0.9) as compared to (0.10) is that we have to minimize the function f which is convex, but so would not be the case when dealing with the gap

$$f^{\#}(x) = \sup\{\langle F(x),x-x'\rangle \mid x' \in Q\}$$

associated with (0.10).

A final comment: solving the applications described above was made possible thanks to the introduction of levels into the bundle approach. In return, the same applications can be solved by the other variants of bundle methods, such as those alluded to in Sections 0.1, 0.2. This may be useful to remove any compactness assumptions; furthermore, the similarity between bundle methods and *sequential quadratic programming* (see [PD 1978]) opens the way to attractive alternatives to the exact penalty approach (cf. the end of Section B above).

In this technical report, we describe the methods and establish their theoretical efficiency estimates. We also give a number of nmerical illustrations (Section 6).

Main notations. $|\cdot|$ denotes the standard Euclidean norm on $\mathbb{R}^n$. If Q is a nonempty closed convex subset in $\mathbb{R}^n$ and $x \in \mathbb{R}^n$, then $\pi(x,Q)$ denotes the (unique) point of Q closest to x.

1. Problems

We consider the following four problems:

(Min) *minimize* $f(x)$ *s.t.* $x \in Q$

Notation and assumptions on the data: f is *convex Lipschitz continuous* on the *bounded closed convex* set $Q \subset \mathbb{R}^n$. L denotes the Lipschitz constant of f, D denotes the diameter of Q with respect to the norm $|\cdot|$ and $V = L\ D$. f^* denotes the minimal value of f on Q.

Oracle: given an input $x \in Q$, computes $f(x)$ and the support functional $f'(x)$ of f at x, $|f'(x)| \le L$.

Accuracy measure: $\varepsilon(x) = \begin{cases} +\infty, & x \notin Q \\ f(x)-\min_Q f, & x \in Q \end{cases}$

(Sad) *find a saddle point of* $f(x,y)$ *on* $Q \times H$

Notation and assumptions on the data: f is *convex in* $x \in Q$, *concave in* $y \in H$ *and Lipschitz continuous* on the direct product of *bounded closed convex* sets $Q \subset \mathbb{R}^n$, $H \subset \mathbb{R}^{n'}$. L_x (L_y) denotes the Lipschitz constant of f with respect to x (resp., y); D_x (D_y) denotes the diameter of Q (resp., H) with respect to the norm $|\cdot|$; V denotes the quantity $L_x\ D_x + L_y\ D_y$.

Oracle: given an input $(x,y) \in Q \times H$, computes $f(x,y)$ and the support functionals $f'_x(x,y)$ of $f(\cdot,y)$ at x and $f'_y(x,y)$ of $f(x,\cdot)$ at y, $|f'_x(x,y)| \le L_x$, $|f'_y(x,y)| \le L_y$.

Accuracy measure: $\varepsilon(x,y) = \begin{cases} +\infty, & (x,y) \notin Q \times H \\ \max_H f(x,\cdot) - \min_Q f(\cdot,y), & (x,y) \in Q \times H \end{cases}$

(CMin) *minimize* $f(x)$ *s.t.* $x \in Q$, $g_i(x) \le 0$, $i = 1,\dots,m$

Notation and assumptions on the data: f is *convex Lipschitz continuous* on the *bounded closed convex* set $Q \subset \mathbb{R}^n$; g_i, $i = 1,\dots,m$, are *convex Lipschitz continuous* on Q. L denotes the maximum of the Lipschitz constants of f, $g_1,\dots,g_m$; D denotes the diameter of Q with respect to the norm $|\cdot|$; $V = DL$, $G = \max\{g_1,\dots,g_m\}$. The problem is assumed to be consistent, and f^* denotes the optimal value of the objective over the feasible set.

Oracle: given an input $x \in Q$, computes $f(x)$, $g_1(x),\dots, g_m(x)$ and the support functionals $f'(x)$, $g'_1(x),\dots, g'_m(x)$ of f, $g_1,\dots,g_m$ at x such that $|f'(x)| \le L$, $|g'_i(x)| \le L$, $i = 1,\dots,m$.

Accuracy measure: $\varepsilon(x) = \begin{cases} +\infty, & x \notin Q \\ \max\{f(x)-f^*, G(x)\}, & x \in Q \end{cases}$

(Var) *find* $x \in Q$ *such that* $F^T(y)(x-y) \ge 0$, $y \in Q$

Notation and assumptions on the data: F is a *monotone bounded-valued operator* on the *bounded closed convex* set $Q \subset \mathbb{R}^n$. L denotes the quantity $\sup_Q |F(\cdot)|$, D denotes the diameter of Q with respect to the norm $|\cdot|$, and V denotes the quantity $L\ D$.

Oracle: given an input $x \in Q$, computes $F(x)$.

Accuracy measure: $\varepsilon(x) = \begin{cases} +\infty, & x \notin Q \\ \max\{F^T(y)(x-y) \mid y \in Q\}, & x \in Q \end{cases}$

2. Methods for (Min)

2.1. Notation. Assume we have called the oracle at the points $x_1,\dots,x_i \in Q$. Then the following objects are defined:

Model: $f_i(x) = \max\{f(x_j) + (f'(x_j))^T(x-x_j) \mid 1 \le j \le i\}$

Remark 2.1.1. Clearly,

$$f_1(x) \le f_2(x) \le \dots \le f_i(x) \le f(x),\ x \in Q, \tag{2.1}$$

all f_j are Lipschitz continuous with Lipschitz constant L and

$$f(x_j) = f_i(x_j),\ 1 \le j \le i. \tag{2.2}$$

ε-subdifferential of the model at $x \in Q$:

$\partial_\varepsilon f_i(x) \equiv \{p \mid f_i(y) \ge f_i(x) - \varepsilon + p^T(y-x)\ \forall\ y \in \mathbb{R}^n\} = \{p = \sum_{j=1}^{i} t_j f'(x_j) \mid t_j \ge 0,\ \sum_{j=1}^{i} t_j = 1,\ \sum_{j=1}^{i} t_j\ \{f(x_j) + (f'(x_j))^T(x_i-x_j)\} \ge f_i(x) - \varepsilon\}$

Remark 2.1.2. From (2.1) - (2.2) it follows immediately that

$$\partial_\varepsilon f_i(x_i) \subset \partial_\varepsilon f(x_i). \tag{2.3}$$

Model's best value: $f_*(i) = \min_Q f_i(\cdot)$

Function's best value: $f^*(i) = \min\{f(x_1),\dots,f(x_i)\}$

Gap: $\Delta(i) = f^*(i) - f_*(i)$

Best point: $x_i^* \in \operatorname{Argmin}\{f(x) \mid x \in \{x_1,\dots,x_i\}\}$

Remark 2.1.3. In view of (2.1) one has

$$\left.\begin{aligned} f_*(1) \le f_*(2) \le \dots \le f_*(i) \le f^* \\ f^*(1) \ge f^*(2) \ge \dots \ge f^*(i) \ge f^* \end{aligned}\right\} \tag{2.4}$$

Remark 2.1.4. In view of (2.4) we have

$$f(x_i^*) - f^* \le \Delta(i) \tag{2.5}$$

and

$$\Delta(1) \ge \Delta(2) \ge \dots \ge \Delta(i) \ge 0 \tag{2.6}$$

Truncated model: $\phi_i(x) = \max\{(f'(x_j))^T(x-x_j) \mid 1 \le j \le i\}$

Remark 2.1.5. Clearly,

$$\phi_1(x) \le \phi_2(x) \le \dots \tag{2.7}$$

and all $\phi_j(\cdot)$ are Lipschitz continuous with Lipschitz constant L.

Truncated model's best value: $\phi_*(i) = \min_Q \phi_i(\cdot)$

Truncated gap: $\delta(i) = -\phi_*(i)$

Remark 2.1.6. The following relations hold:

$$\delta(1) \ge \delta(2) \ge \dots \ge \delta(i) \ge 0 \tag{2.8}$$

$$\phi_i(x_i) \ge 0; \; f(x_i^*) - f^* \le \delta(i). \tag{2.9}$$

□ Monotonicity of $\delta(\cdot)$ immediately follows from (2.7). To prove nonnegativity of $\delta(i)$, let x^* be an optimal solution to (Min). Then $(f'(x_j))^T(x^*-x_j) \le 0$ for all j, so that $\phi_i(x^*) \le 0$. (2.8) is proved. The first relation in (2.9) is evident. To prove the second relation, note that $f(x^*) \ge f(x_j) + (f'(x_j))^T(x^*-x_j) \ge f(x_i^*) + (f'(x_j))^T(x^*-x_j)$, $j = 1,\dots, i$, whence $f(x^*) \ge f(x_i^*) + \phi_i(x^*) \ge f(x_i^*) + \phi_*(i)$. ■

2.2. Methods

2.2.1. Level Method (LM)

A. Description of LM

Parameters: $\lambda \in (0,1)$

Initialization: x_1 is an arbitrary point of Q

i-th step:

1) Call the oracle, x_i being the input

2) Compute $f_*(i)$, $f^*(i)$, $x^*(i)$

3) Set

$$l(i) = f_*(i) + \lambda\, \Delta(i),$$

$$x_{i+1} = \pi(x_i, \{x \mid x \in Q,\; f_i(x) \le l(i)\})$$

B. Efficiency estimate. We claim that

$$\varepsilon(x_i^*) \le \Delta(i),$$
$$i > c(\lambda)\ (V/\varepsilon)^2 \Rightarrow \varepsilon(x_i^*) \le \varepsilon,$$

where

$$c(\lambda) = (1-\lambda)^{-2}\ \lambda^{-1}\ (2-\lambda)^{-1}$$

(note that $\min\ c(\cdot) = 4 = c(0.29289...)$).

Proof.

B.1. The efficiency estimate

$$\varepsilon(x_i^*) \le \Delta(i) \qquad \text{(LM.1)}$$

was established in (2.5).

B.2. Set $S_i = [f_*(i), f^*(i)]$. Then (see (2.4))

$$S_1 \supseteq S_2 \supseteq \dots,\ |S_i| = \Delta(i), \qquad \text{(LM.2)}$$

where $|S|$ denotes the length of the segment S.

B.3. Lemma. Let $i'' > i'$ be such that

$$\Delta(i'') \ge (1-\lambda)\ \Delta(i'). \qquad \text{(LM.3)}$$

Then

$$f_*(i'') \le l(i'). \qquad \text{(LM.4)}$$

□ Indeed, the length of the segment $\{s \in S_{i'} \mid s \ge l(i)\}$ is $(1-\lambda)\ \Delta(i')$ and, since $S_{i'} \supseteq S_{i''}$ ((LM.1)), the converse of (LM.4) would imply $\Delta(i'') = |S_{i''}| < (1-\lambda)\ \Delta(i')$, which is impossible. ■

B.4. Let us fix $\varepsilon > 0$ and assume that for certain N and all $i \le N$ we have $\Delta(i) > \varepsilon$. Let us split the integer segment $I = 1,\dots,N$ in groups $I_1,\dots, I_k$ as follows. The last element of the first group is $j_1 \equiv N$, and this group contains precisely those $i \in I$ for which $\Delta(i) \le (1-\lambda)^{-1}\Delta(j_1)$. The largest element of I, j_2, which does not belong to the group I_1, if such an element exists, is the last element of I_2, and the latter group consists precisely of those $i \le j_2$, for which $\Delta(i) \le (1-\lambda)^{-1}\ \Delta(j_2)$. The largest element of I, j_3, which does not belong to I_2, is the last element of I_3, and this group consists of those $i \le j_3$ satisfying $\Delta(i) \le (1-\lambda)^{-1} \Delta(j_3)$, and so on. Let $u(l)$ be the minimizer of the function $f_{j_l}(\cdot)$ over Q. Lemma B.3, applied with an arbitrary $i' \in I_l$ and with $i'' = j_l$, demonstrates that $f_*(j_l)) = f_{j_l}(u(l)) \le l(i)$ for all $i \in I_l$. (2.1) shows that $f_j(u(l)) \le l(i)$ for all $i,j \in I_l$. Thus, we have established the following:

the (clearly convex) level sets $Q_i = \{x \in Q \mid f_i(x) \le l(i)\}$ associated with $i \in I_l$, have a common point (namely, $u(l)$). (LM.5)

B.5. By virtue of standard properties of the projection mapping, (LM.5) imply

$$\tau_{i+1} \equiv |x_{i+1}-u(l)|^2 \le \tau_i - \text{dist}^2\{x_i,Q_i\},\ i \in I_l. \qquad \text{(LM.6)}$$

We also have $f_i(x_i) - l(i) = f(x_i) - l(i) \ge f^*(i) - l(i) = (1-\lambda)\Delta(i)$ and $f_i(x_{i+1}) \le l(i)$. From the Lipschitz property of f_i, it follows that $\text{dist}\{x_i,Q_i\} = |x_i - x_{i+1}| \ge L^{-1}|f_i(x_i) - f_i(x_{i+1})| \ge L^{-1} (1-\lambda)\Delta(i)$. Thus,

$$\tau_{i+1} \le \tau_i - L^{-2} (1-\lambda)^2 \Delta^2(i) \le \tau_i - L^{-2}(1-\lambda)^2\Delta^2(j_l),\ i \in I_l.$$

Because $0 \le \tau_i \le D^2$ (evident), the latter inequality immediately implies that the number N_l of elements in I_l satisfies the estimate

$$N_l \le D^2L^2(1-\lambda)^{-2}\Delta^{-2}(j_l). \qquad \text{(LM.7)}$$

B.6. Form the definitions of N and of a group, we have

$$\Delta(j_1) = \Delta(N) > \varepsilon,\ \Delta(j_{l+1}) > (1-\lambda)^{-1}\Delta(j_l).$$

These relations combined with (LM.7) imply $N = \sum_{l\ge 1} N_l \le D^2L^2(1-\lambda)^{-2} \sum_{l\ge 1} \varepsilon^{-2} (1-\lambda)^{2(l-1)} = (V/\varepsilon)^2 (1-\lambda)^{-2} \lambda^{-1} (2-\lambda)^{-1}$. ∎

2.2.2. Proximal Level Method (PLM)

A. Description of PLM

Parameters: $\lambda \in (0,1)$; $\mu = (1-\lambda)$

Initialization: x_1 is an arbitrary point of Q; $\Delta'(0) = \infty$

***i*-th step:**

1) Call the oracle, x_i being the input

2) Compute $f_*(i)$, $f^*(i)$, $x^*(i)$

3) Set

$$l(i) = f_*(i) + \lambda\ \Delta(i),$$

$$l'(i) = \begin{cases} l(i), \text{ if } \Delta(i) < \mu\Delta'(i-1) \\ \min\{l(i), l'(i-1)\}, \text{ otherwise} \end{cases}$$

$$\Delta'(i) = \begin{cases} \Delta(i), \text{ if } \Delta(i) < \mu\Delta'(i-1) \\ \Delta'(i-1), \text{ otherwise} \end{cases}$$

$$x_{i+1} = \pi(x_i^*, \{x \mid x \in Q,\ f_i(x) \le l'(i)\})$$

Remark. The difference between PLM and LM is first that, in PLM, x_{i+1} is the projection of the i-th best point x_i^* (and not the i-th iterate x_i) onto the level set of the i-th model f_i; second, the levels defining the above level sets are different: in LM this quantity, $l(i)$, divides in a fixed ratio the segment $[f_*(i), f^*(i)]$, and it can increase as well as decrease, as i varies, while in PLM the corresponding quantity is forbidden to increase until the gap $f^*(i) - f_*(i)$ decreases "substantially".

B. Efficiency estimate. We claim that

$$\varepsilon(x_i^*) \le \Delta(i),$$

$$i > c(\lambda)\ (V/\varepsilon)^2 \Rightarrow \varepsilon(x_i^*) \le \varepsilon,$$

$$c(\lambda) = (1-\lambda)^{-4}\ (2-\lambda)^{-1}\ \lambda^{-1}$$

(note that $\min c(\cdot) = 6.75 = c(0.18350...)$).

Proof.

B.1. The efficiency estimate

$$\varepsilon(x_i^*) \le \Delta(i) \qquad \text{(PLM.1)}$$

was established in (2.5).

B.2. Set $S_i = [f_*(i), f^*(i)]$. Then (see (2.4))

$$S_1 \supseteq S_2 \supseteq \dots, \ |S_i| = \Delta(i), \qquad \text{(PLM.2)}$$

where $|S|$ denotes the length of a segment S.

B.3. Let us fix $\varepsilon > 0$ and assume that for certain N and all $i \le N$ we have $\Delta(i) > \varepsilon$. Let us split the integer segment $I = 1,\dots,N$ in groups $I_1,\dots, I_k$ as follows. The first element of the first group is $i_1 = 1$, and this group contains precisely those $i \in I$ for which $\Delta(i) \ge \mu\Delta(i_1)$. The smallest element of I, i_2, which does not belong to the group I_1, if such an element exists, is the first element of I_2, and the latter group consists precisely of those $i \ge i_2$, for which $\Delta(i) \ge \mu\Delta(i_2)$. The smallest element of I, i_3, which does not belong to $I_1 \cup I_2$, is the first element of I_3, and this group consists of those $i \ge i_3$ satisfying $\Delta(i) \ge \mu\Delta(i_3)$, and so on. Note that the following relations come from the description of the method:

$$\Delta'(i) = \Delta(i_l), \ i \in I_l; \qquad \text{(PLM.3)}$$

$$l'(i_l) = l(i_l), \ l'(i) = \min\{l'(i-1), l(i)\}, \ i \in I_l \backslash \{i_l\}. \qquad \text{(PLM.4)}$$

Lemma 2.2.1.B.3 implies that, for all i', $i'' \in I_l$, $i' \le i''$, we have $f_*(i'') \le l(i')$. Combined with (PLM.4), this observation means that $f_*(i'') \le l'(i')$ if i', $i'' \in I_l$ and $i' \le i''$. In particular, the level sets $Q_i = \{x \in Q \mid f_i(x) \le l'(i)\}$ are nonempty, so that the method is well-defined.

Now note that $Q_i \supseteq Q_{i+1}$, if i, $i+1 \in I_l$, since $f_{i+1}(\cdot) \ge$

$f_i(\cdot)$ and $l'(i+1) \le l'(i)$. Thus,

the (clearly convex) level sets $Q_i = \{x \in Q \mid f_i(x) \le l'(i)\}$ *associated with* $i \in I_l$ *are nonempty and contain* Q_{j_l}. (PLM.5)

B.5. For a fixed l, let us divide the group I_l into the sequential subgroups $J_1,\dots,J_q$ in such a way that the best points x_i^* associated with $i \in J_r$ coincide with each other and differ from the remaining best points associated with other $i \in I_l$. Thus, $x_i^* = x(r)$ for $i \in J_r$, and the points $x(1),\dots,x(q)$ are mutually different. In view of the description of the method we have

$$\left.\begin{aligned} x_{i+1} &= \pi(x(r),Q_i), \ i \in J_r, \\ x(r+1) &= x_{i(r)+1} = \pi(x(r),Q_{i(r)}), \text{ if } J_{r+1} \ne \varnothing, \end{aligned}\right\} \qquad \text{(PLM.6)}$$

where $i(r)$ is the last element of J_r.

B.6. By virtue of the standard properties of the projection mapping, from the inclusions $Q_i \subseteq Q_{i-1}$ it follows for for $i \in J_r$:

$$\tau_{i+1} \equiv |x(r)-x_{i+1}|^2 \ge \tau_i + |x_i-x_{i+1}|^2. \qquad \text{(PLM.7)}$$

We also have $f_i(x(r)) - l'(i) = f^*(i) - l'(i) \ge f^*(i) - l(i) = (1 - \lambda)\,\Delta(i) > 0$, so that $x(r)$ does not belong to Q_i; it immediately follows that $f_i(x_{i+1}) = l'(i) \le l(i)$, while $f_i(x_i) = f(x_i) \ge f^*(i) \ge (1-\lambda)\ \Delta(i) + l(i)$. Thus, $f_i(x_i) - f_i(x_{i+1}) \ge (1-\lambda)\Delta(i)$, and since f_i is Lipschitz continuous with constant L, we conclude that $|x_i-x_{i+1}| \ge L^{-1}(1-\lambda)\Delta(i) \ge L^{-1}(1-\lambda)^2\Delta(i_l)$. This inequality, combined with (PLM.7), means that $|x(r) - x_{i(r)+1}|^2 \ge |J_r|L^{-2}(1-\lambda)^4 \Delta^2(i_l)$, where $|J_r|$ denotes the cardinality of J_r.

Now let us minorize the quantity $R^2 = |x(1) - x_{j_l+1}|^2$, where j_l is the last element of I_l. We have: $x(1)$ is a certain point of Q; $x(2)$ is the projection of $x(1)$ onto $Q_{i(1)}$; $x(3)$ is the projection of $x(2)$ onto $Q_{i(2)}, \ldots,$ $x(q)$ is the projection of $x(q-1)$ onto $Q_{i(q-1)}$, and x_{j_l+1} is the projection of $x(q)$ onto Q_{j_l}. The sets $Q_{\bullet}$ involved in the latter family are included the next into the previous, so that $R^2 \geq |x(1) - x(2)|^2 + \ldots + |x(q-1) - x(q)|^2 + |x(q) - x_{j_l+1}|^2$; the latter sum, as it was proved, is not less than $L^{-2}(1-\lambda)^4\Delta^2(i_l) \sum_r |J_r| = L^{-2}(1-\lambda)^4\Delta^2(i_l)\ |I_l|$. On the other hand, we clearly have $R^2 \leq D^2$, whence

$$|I_l| \leq L^2\ D^2\ (1-\lambda)^{-4}\ \Delta^{-2}(i_l). \tag{PLM.8}$$

We have $\Delta(i_k) > \varepsilon$ (k is is the number of the last group I_l in the segment $I = 1,\ldots,N$) and $\Delta(i_{l-1}) > (1-\lambda)^{-1}\ \Delta(i_l)$ (the latter inequality is a consequence of our definition of the groups $I_{\bullet}$). Therefore $N = \sum_{l=1}^{k} |I_l| = \sum_{l=1}^{k} |I_{k+1-l}| \leq L^2\ D^2\ (1-\lambda)^{-4}\ \Delta^{-2}(i_k) \sum_{l=1}^{k} (1-\lambda)^{2(l-1)} \leq L^2\ D^2\ (1-\lambda)^{-4}\ \varepsilon^{-2}\ (2-\lambda)^{-1}\ \lambda^{-1} = c(\lambda)\ (V/\varepsilon)^2$. ■

2.2.3. Dual Level Method (DLM)

A. Description of DLM

Parameters: $\lambda, \mu \in (0,1)$

Initialization: x_1 is an arbitrary point of Q

***i*-th step:**

1) Call the oracle, x_i being the input

2) Compute $f^*(i),\ f_*(i),\ x_i^*$

3) Set

$$l(i) = f_*(i) + \lambda\ \Delta(i)\ (= f^*(i) - (1-\lambda)\ \Delta(i)),$$

$$\varepsilon^+(i) = f(x_i) - l(i) - \mu\ (1-\lambda)\ \Delta(i)$$

(note that $\varepsilon^+(i) \geq 0$, since $f(x_i) - l(i) \geq f^*(i) - l(i) = (1-\lambda)\ \Delta(i)$). Define p_i as the solution to the problem

$P(i)$: $\quad$ *minimize* $|p|^2$ *subject to* $p \in \partial_{\varepsilon^+(i)} f_i(x_i)$

and set

$$x_{i+1} = \pi(x_i - \mu(1-\lambda)\Delta(i)\ |p_i|^{-2}\ p_i, Q).$$

B. Efficiency estimate. We claim that

$$\varepsilon(x_i^*) \le \Delta(i),$$

$$i > c(\lambda,\mu)\ (V/\varepsilon)^2 \Rightarrow \varepsilon(x_i^*) \le \varepsilon,$$

$$c(\lambda,\mu) = \mu^{-2}(1-\lambda)^{-2}\ \lambda^{-1}\ (2-\lambda)^{-1}$$

(note that $\min_\lambda c(\lambda,\mu) = 4\ \mu^{-2} = c(0.29289...,\mu)$).

Proof.

B.1. The efficiency estimate

$$\varepsilon(x_i^*) \le \Delta(i) \tag{DLM.1}$$

was established in (2.5).

B.2. Set $S_i = [f_*(i), f^*(i)]$. Then (see (2.4))

$$S_1 \supseteq S_2 \supseteq \ldots,\ |S_i| = \Delta(i), \tag{DLM.2}$$

where $|S|$ denotes the length of a segment S.

B.3. Let us fix $\varepsilon > 0$ and assume that for certain N and all $i \le N$ we have $\Delta(i) > \varepsilon$. Let us split the integer segment $I = 1,...,N$ in groups $I_1,..., I_k$ as follows. The last element of the first group is $j_1 \equiv N$, and this group contains precisely those $i \in I$ for which $\Delta(i) \le (1-\lambda)^{-1}\Delta(j_1)$. The largest element of I, j_2, which does not belong to the group I_1, if such an element exists, is the last element of I_2, and the latter group consists precisely of those $i \le j_2$, for which $\Delta(i) \le (1-\lambda)^{-1}\ \Delta(j_2)$. The largest element of I, j_3, which does not belong to I_2, is the last element of I_3, and this group consists of those $i \le j_3$ satisfying $\Delta(i) \le (1-\lambda)^{-1}\ \Delta(j_3)$, and so on.

Let $u(l)$ be the minimizer of the function $f_{j_l}(\cdot)$ over Q. Lemma 2.2.1.B.3, applied with an arbitrary $i' \in I_l$ and $i'' = j_l$, demonstrates that $f_*(j_l)) = f_{j_l}(u(l)) \le l(i)$ for all $i \in I_l$. (2.1) shows that $f_j(u(l)) \le l(i)$ for all $i,j \in I_l$. Thus, we have established the following:

the (clearly convex) level sets $Q_i = \{x \in Q \mid f_i(x) \le l(i)\}$ *associated with* $i \in I_l$, *have a common point (namely,* $u(l)$*).* (DLM.3)

B.4. Let $i \in I_l$. The definition of p_i implies $f^i(x) \equiv f(x_i) + p_i^T (x-x_i) - \varepsilon^+(i) \le f_i(x)$. In particular, $f^i(u(l)) \le f_i(u(l)) \le l(i)$, while $f^i(x_i) = f(x_i) - \varepsilon^+(i) = l(i) + \mu (1-\lambda) \Delta(i) \ge l(i)$. We conclude that $f^i(x_i) - f^i(u(l)) \ge \mu(1-\lambda)\Delta(i)$, so that $p_i^T(x_i-u(l)) \ge \mu(1-\lambda)\Delta(i)$. Since $x_{i+1} = \pi(x_i - \mu(1-\lambda)\Delta(i) |p_i|^{-2} p_i, Q)$, it follows that

$$\tau_{i+1} \equiv |x_{i+1} - u(l)|^2 \le \tau_i - |p_i|^{-2} \mu^2(1-\lambda)^2\Delta^2(i);$$

clearly, $|p_i| \le L$ (since f_i is Lipschitz continuous with constant L) and $\Delta(i) \ge \Delta(j_l)$, and we obtain

$$\tau_{i+1} \equiv |x_{i+1} - u(l)|^2 \le \tau_i - L^{-2} \mu^2(1-\lambda)^2\Delta^2(j_l),\ i \in I_l. \qquad \text{(DLM.4)}$$

Because $0 \le \tau_i \le D^2$ (evident), the latter inequality immediately implies that the number N_l of elements in I_l satisfies the estimate

$$N_l \le D^2L^2\mu^{-2}(1-\lambda)^{-2}\Delta^{-2}(j_l). \qquad \text{(DLM.5)}$$

B.5. From the definitions of N and of a group, we have

$$\Delta(j_1) = \Delta(N) > \varepsilon,\ \Delta(j_{l+1}) > (1-\lambda)^{-1}\Delta(j_l).$$

These relations combined with (DLM.5) imply $N = \sum_{l\ge1} N_l \le D^2L^2\mu^{-2}(1-\lambda)^{-2} \sum_{l\ge1} \varepsilon^{-2} (1-\lambda)^{2(l-1)} = (V/\varepsilon)^2\mu^{-2}(1-\lambda)^{-2} \lambda^{-1} (2-\lambda)^{-1}$. ■

2.2.4. Truncated Level Method (TLM)

A. Description of TLM

Parameters: $\lambda \in (0,1)$

Initialization: x_1 is an arbitrary point of Q

***i*-th step:**

1) Call the oracle, x_i being the input

2) Compute $\phi_*(i)$, $f^*(i)$, x_i^*

3) Set

$$l(i) = -(1-\lambda)\ \delta(i),$$

$$x_{i+1} = \pi(x_i,\{x \mid x \in Q,\ \phi_i(x) \le l(i)\}) .$$

Remark. The difference between LM and TLM is that the latter method uses an artificial model which involves only subgradients, not the values of the objective. This feature of TLM is not valuable in the case of (Min), but it will be useful for (Var).

B. Efficiency estimate. We claim that

$$\varepsilon(x_i^*) \le \delta(i),$$

$$i > c(\lambda)\ (V/\varepsilon)^2 \Rightarrow \varepsilon(x_i^*) \le \varepsilon,$$

$$c(\lambda) = (1-\lambda)^{-2}\ \lambda^{-1}\ (2-\lambda)^{-1}$$

(note that $\min c(\cdot) = 4 = c(0.29289...)$).

Proof.

B.1. The efficiency estimate

$$\varepsilon(x_i^*) \le \delta(i) \tag{TLM.1}$$

was established in (2.9).

B.2. Set $S_i = [\phi_*(i),0]$. Then (see (2.7), (2.8)) $S_i \ne \emptyset$ and

$$S_1 \supseteq S_2 \supseteq ...,\ |S_i| = \delta(i), \tag{TLM.2}$$

where $|S|$ denotes the length of a segment S.

B.3. Let us fix $\varepsilon > 0$ and assume that for certain N and all i

$\leq N$ we have $\delta(i) > \varepsilon$. Let us split the integer segment $I = 1,\dots,N$ in groups $I_1,\dots, I_k$ as follows. The last element of the first group is $j_1 \equiv N$, and this group contains precisely those $i \in I$ for which $\delta(i) \leq (1-\lambda)^{-1}\delta(j_1)$. The largest element of I, j_2, which does not belong to the group I_1, if such an element exists, is the last element of I_2, and the latter group consists precisely of those $i \leq j_2$, for which $\delta(i) \leq (1-\lambda)^{-1}\ \delta(j_2)$. The largest element of I, j_3, which does not belong to I_2, is the last element of I_3, and this group consists of those $i \leq j_3$ satisfying $\delta(i) \leq (1-\lambda)^{-1}$ $\delta(j_3)$, and so on.

B.4. From (TLM.2) it immediately follows that $\phi_*(j_l) \leq l(i)$, $i \in I_l$. Let $u(l)$ be the minimizer of the function $\phi_{j_l}(\cdot)$ over Q; then for $i \in I_l$ one has $\phi_i(u(l)) \leq \phi_{j_l}(u(l)) \leq l(i)$. Thus, we have established that

the (clearly convex) level sets $Q_i = \{x \in Q \mid \phi_i(x) \leq l(i)\}$ *associated with* $i \in I_l$, *have a common point (namely,* $u(l)$*).* (TLM.3)

B.5. The standard properties of the projection mapping and (TLM.3) imply

$$\tau_{i+1} \equiv |x_{i+1}-u(l)|^2 \leq \tau_i - \mathrm{dist}^2\{x_i,Q_i\},\ i \in I_l. \qquad \text{(TLM.4)}$$

We also have $\phi_i(x_i) - l(i) \geq -\,l(i)$ (see (2.9)), so that $\phi_i(x_i) - l(i) \geq (1-\lambda)\delta(i)$, and $\phi_i(x_{i+1}) \leq l(i)$. Since ϕ_i is Lipschitz continuous with the constant L, it follows that $\mathrm{dist}\{x_i,Q_i\} = |x_i - x_{i+1}| \geq L^{-1}|\phi_i(x_i) - \phi_i(x_{i+1})| \geq L^{-1}\ (1-\lambda)\delta(i)$. Thus,

$$\tau_{i+1} \leq \tau_i - L^{-2}\ (1-\lambda)^2\ \delta^2(i) \leq \tau_i - L^{-2}(1-\lambda)^2\delta^2(j_l),\ i \in I_l.$$

Because $0 \leq \tau_i \leq D^2$ (evident), the latter inequality immediately implies that the number N_l of elements in I_l satisfies the estimate

$$N_l \le D^2 L^2 (1-\lambda)^{-2} \delta^{-2}(j_l). \qquad \text{(TLM.5)}$$

B.6. From the definitions of N and of a group, we have

$$\delta(j_1) = \delta(N) > \varepsilon, \ \delta(j_{l+1}) > (1-\lambda)^{-1}\delta(j_l).$$

These relations combined with (TLM.5) imply $N = \sum_{l\ge 1} N_l \le D^2 L^2 (1-\lambda)^{-2}$ $\sum_{l\ge 1} \varepsilon^{-2} (1-\lambda)^{2(l-1)} = (V/\varepsilon)^2 (1-\lambda)^{-2} \lambda^{-1} (2-\lambda)^{-1}$. ∎

3. Methods for (Sad)

3.0. Initial scaling. *In what follows we assume that the diameters of Q and H coincide; D denotes their (common) value.* This assumption can be provided by an appropriate isotropic scaling of, say, the y-variable. Note that the quantity $L_x D_x + L_y D_y$ remains invariant under this scaling. We denote $L = \max\{L_x, L_y\}$.

3.1. Notation. Denote

$$\overline{f}(x) = \max_H f(x,\cdot): Q \to \mathbb{R}, \qquad \underline{f}(y) = \min_Q f(\cdot,y): H \to \mathbb{R}$$

(these are, respectively, the worst-case payment of the player choosing x and the worst-case income of the player choosing y in the game associated with f).

Assume we have called the oracle at the points $(x_1,y_1),\dots,(x_i,y_i) \in Q$. Then the following objects are defined:

Models:

x-model:

$$\overline{f}_i(x) = \max\{f(x_j,y_j) + (f'_x(x_j,y_j))^T(x-x_j) \mid 1 \le j \le i\}: Q \to \mathbb{R},$$

y-model:

$$\underline{f}_i(y) = \min\{f(x_j,y_j) + (f'_y(x_j,y_j))^T(y-y_j) \mid 1 \le j \le i\}: H \to \mathbb{R},$$

model:

$$f_i(x,y) = \overline{f}_i(x) - \underline{f}_i(y): Q\times H \to \mathbb{R}.$$

Remark 3.1.1. Clearly, $\overline{f}_i$ is convex, $\underline{f}_i$ is concave,

$$\overline{f}_1(x) \le \overline{f}_2(x) \le \dots \le \overline{f}_i(x) \le \overline{f}(x),\ x \in Q, \tag{3.1}$$

$$\underline{f}_1(y) \ge \underline{f}_2(y) \ge \dots \ge \underline{f}_i(y) \ge \underline{f}(y),\ y \in H, \tag{3.2}$$

$\overline{f}_i$, $\underline{f}_i$ are Lipschitz continuous with Lipschitz constant L. Consequently,

$$f_1(x,y) \le f_2(x,y) \le \dots \le f_i(x,y) \le \overline{f}(x)-\underline{f}(y),\ (x,y) \in Q\times H, \tag{3.3}$$

and f_i is Lipschitz continuous with Lipschitz constant $2^{1/2}L$.

ε-subdifferential of the model at $x \in Q$:

$$\partial_\varepsilon f_i(x,y) \equiv \{p \in \mathbb{R}^n\times\mathbb{R}^{n'} \mid f_i(u,v) \ge f_i(x,y) - \varepsilon + p^T((u,v)-(x,y))\ \forall\ (u,v) \in \mathbb{R}^n\times\mathbb{R}^{n'}\}$$

Model's best value: $f_*(i) = \min_{Q\times H} f_i(\cdot,\cdot)$

Gap: $\Delta(i) = -f_*(i)$

Remark 3.1.2. The following relations hold:

$$\Delta(1) \ge \Delta(2) \ge \dots \ge \Delta(i) \ge 0;\ f_i(x_i,y_i) \ge 0. \tag{3.4}$$

□ Indeed, the monotonicity of $\Delta(\cdot)$ follows from (3.3). Let us prove that $\Delta(\cdot)$ is nonnegative. Let $f^* = \min_Q \overline{f}(\cdot)$; by von Neumann's lemma, one also has $f^* = \max_H \underline{f}(\cdot)$. It follows that $\min_{Q\times H} (\overline{f}(x) - \underline{f}(y)) = f^* - f^* = 0$, and the first relation in (3.4) follows from (3.3). On the other hand, clearly $\overline{f}_i(x_i) \ge f(x_i,y_i)$, $\underline{f}_i(y_i) \le f(x_i,y_i)$, which implies the second relation in (3.4). ■

Truncated model:

$$\phi_i(x,y) = \max\{(f'_x(x_j,y_j))^T(x-x_j) - (f'_y(x_j,y_j))^T(y-y_j) \mid 1 \le j \le i\}: Q\times H \to \mathbb{R}.$$

Remark 3.1.3. Clearly, $\phi_i(x,y)$ is convex and Lipschitz continuous with Lipschitz constant $2^{1/2}L$, and

$$\phi_1(\cdot,\cdot) \le \phi_2(\cdot,\cdot) \le \dots \tag{3.5}$$

Truncated model's best value: $\phi_*(i) = \min_{Q\times H} \phi_i(\cdot,\cdot)$

Truncated gap: $\delta(i) = -\phi_*(i)$.

Remark 3.1.4. We have

$$\delta(1) \geq \delta(2) \geq \ldots \geq 0 \tag{3.6}$$

□ The monotonicity of $\delta(\cdot)$ follows from (3.5). Let us prove that $\delta(i) \leq 0$. Indeed, let (x^*,y^*) be a saddle point of f and let $(x,y) \in Q\times H$. We have $f(x^*,y) \geq (f'_x(x,y))^T(x^*-x) + f(x,y)$, $f(x,y^*) \leq (f'_y(x,y))^T(y^*-y) + f(x,y)$, whence $f(x^*,y) - f(x,y^*) \geq (f'_x(x,y))^T (x^*-x) - (f'_y(x,y))^T(y^*-y)$. Since (x^*,y^*) is saddle point, $f(x^*,y) - f(x,y^*) \leq 0$, so that $(f'_x(x,y))^T(x^*-x) - (f'_y(x,y))^T(y^*-y) \leq 0$, $(x,y) \in Q\times H$. In other words, $\phi_i(x^*,y^*) \leq 0$. ■

3.2. Methods

3.2.1. Level Method (LM)

A. Description of LM

Parameters: $\lambda \in (0,1)$

Initialization: (x_1,y_1) is an arbitrary point of $Q\times H$

***i*-th step:**

1) Call the oracle, (x_i,y_i) being the input

2) Compute $f_*(i)$, i.e., solve the pair of convex problems

$P_x(i)$: *minimize*

$$\bar{f}_i(x) = \max\{f(x_j,y_j) + (f'_x(x_j,y_j))^T(x-x_j) \mid 1 \leq j \leq i\}$$

subject to $x \in Q$

and

$P_y(i)$: *maximize*

$$\underline{f}_i(y) = \max\{f(x_j,y_j) + (f'_y(x_j,y_j))^T(y-y_j) \mid 1 \leq j \leq i\}$$

subject to $y \in H$.

3) Set

$$l(i) = f_*(i) + \lambda\, \Delta(i),$$

$$(x_{i+1}, y_{i+1}) = \pi((x_i, y_i), \{(x,y) \mid (x,y) \in Q\times H,\ f_i(x,y) \le l(i)\}).$$

The i-th approximate solution is defined as follows. When solving the problems $P_x(i)$ and $P_y(i)$, we find also optimal dual solutions, i.e., the quantities $\{t_i(j), s_i(j)\}_{1\le j\le i}$, satisfying

$$\sum_{j=1}^{i} t_i(j) = 1,\ t_i(j) \ge 0,\ \min_{x\in Q} \sum_{j=1}^{i} t_i(j)\{f(x_j,y_j) + (f'_x(x_j,y_j))^T (x - x_j)\} = \min_Q \overline{f}_i(\cdot),$$

$$\sum_{j=1}^{i} s_i(j) = 1,\ s_i(j) \ge 0,\ \max_{y\in H} \sum_{j=1}^{i} s_i(j)\{f(x_j,y_j) + (f'_y(x_j,y_j))^T (y - y_j)\} = \max_H \underline{f}_i(\cdot),$$

and the i-th approximate solution is defined as

$$(x_i^* = \sum_{j=1}^{i} s_i(j)\, x_j,\ y_i^* = \sum_{j=1}^{i} t_i(j)\, y_j).$$

B. Efficiency estimate. We claim that

$$\varepsilon(x_i^*, y_i^*) \le \Delta(i),$$

$$i > c(\lambda)\, (V/\varepsilon)^2 \Rightarrow \varepsilon(x_i^*, y_i^*) \le \varepsilon,$$

where

$$c(\lambda) = 4\, (1-\lambda)^{-2}\, \lambda^{-1}\, (2-\lambda)^{-1}$$

(note that $\min c(\cdot) = 16 = c(0.29289...)$).

Proof.

B.1. Let us fix $(x,y) \in Q\times H$. We have

$$f(x_j,y) \le f(x_j,y_j) + (f'_y(x_j,y_j))^T (y-y_j),$$

$$f(x,y_j) \ge f(x_j,y_j) + (f'_x(x_j,y_j))^T (x-x_j).$$

It follows that $\sum_{j=1}^{i} t_i(j)\, f(x,y_j) - \sum_{j=1}^{i} s_i(j)\, f(x_j,y) \ge \sum_{j=1}^{i} t_i(j)$ $(f(x_j,y_j) + (f'_x(x_j,y_j))^T (x-x_j)) - \sum_{j=1}^{i} s_i(j)\ (f(x_j,y_j) + (f'_y(x_j,y_j))^T (y-y_j))$. Since f is convex in x and concave in y, we have $\sum_{j=1}^{i} t_i(j)\ f(x,y_j) - \sum_{j=1}^{i} s_i(j)\ f(x_j,y) \le f(x,y_i^*) - f(x_i^*,y)$.

Thus,

$$f(x,y_i^*) - f(x_i^*,y) \ge \sum_{j=1}^{i} t_i(j)\ (f(x_j,y_j) + (f'_x(x_j,y_j))^T(x-x_j)) -$$
$$- \sum_{j=1}^{i} s_i(j)\ (f(x_j,y_j) + (f'_y(x_j,y_j))^T(y-y_j)),\ (x,y) \in Q\times H.$$

Taking the minimum over $(x,y) \in Q\times H$ and using the definition of $t_i(\cdot)$, $s_i(\cdot)$, we obtain

$$\overline{f}(x_i^*) - \underline{f}(y_i^*) \le \max_H \underline{f}_i(\cdot) - \min_Q \overline{f}_i(\cdot) = - \min_{Q\times H} f_i(x,y).$$

In other words,

$$\varepsilon(x_i^*,y_i^*) \le \Delta(i), \qquad \text{(LM.1)}$$

as is required in the accuracy estimate.

B.2. Set $S_i = [\phi_*(i),0]$. Then (see (3.3), (3.4)) $S_i \ne \emptyset$ and

$$S_1 \supseteq S_2 \supseteq \ldots,\ |S_i| = \Delta(i), \qquad \text{(LM.2)}$$

where $|S|$ for a segment S denotes the length of S.

B.3. Let us fix $\varepsilon > 0$ and assume that for certain N and all $i \le N$ we have $\delta(i) > \varepsilon$. Let us split the integer segment $I = 1,\ldots,N$ in groups $I_1,\ldots, I_k$ as follows. The last element of the first group is $j_1 \equiv N$, and this group contains precisely those $i \in I$ for which $\delta(i) \le (1-\lambda)^{-1}\delta(j_1)$. The largest element of I, j_2, which does not belong to the group I_1, if such an element exists, is the last element of I_2, and the latter group consists precisely of those $i \le j_2$, for which $\delta(i) \le (1-\lambda)^{-1}\ \delta(j_2)$. The largest element of I, j_3, which does not belong to I_2, is the last element of I_3, and this group consists of those $i \le j_3$ satisfying $\delta(i) \le (1-\lambda)^{-1}$ $\delta(j_3)$, and so on.

Let $(u(l),v(l))$ minimize the function $f_{j_l}(\cdot,\cdot)$ over $Q\times H$. For $i \in I_l$ from (LM.2), the definition of $l(i)$ and the relation $\delta(j_l)$

$\geq (1-\lambda)\ \delta(i),\ i \in I_l$, it immediately follows that $f_*(j_l)) = f_{j_l}(u(l),v(l)) \leq l(i)$ for all $i \in I_l$. (3.3) shows that $f_j(u(l),v(l)) \leq l(i)$ for all $i,j \in I_l$. Thus, we have established the following:

the (clearly convex) level sets $Q_i = \{z \in Q\times H \mid f_i(z) \leq l(i)\}$ *associated with* $i \in I_l$, *have a common point (namely,* $z(l) = (u(l), v(l)))$. (LM.3)

B.4. By virtue of the standard properties of the projection mapping, (LM.3), under the notation $z_i = (x_i,y_i)$, implies

$$\tau_{i+1} \equiv |z_{i+1}-z(l)|^2 \leq \tau_i - \mathrm{dist}^2\{z_i,Q_i\},\ i \in I_l. \tag{LM.4}$$

We also have $f_i(z_i) - l(i) \geq - l(i)$ (see (3.4)), whence $f_i(z_i) - l(i) \geq (1-\lambda)\delta(i)$, while $f_i(z_{i+1}) \leq l(i)$. Since f_i is Lipschitz continuous with the constant $2^{1/2}L$, it follows that $\mathrm{dist}\{z_i,Q_i\} = |z_i - z_{i+1}| \geq 2^{-1/2}L^{-1}|f_i(z_i) - f_i(z_{i+1})| \geq 2^{-1/2}L^{-1}\ (1-\lambda)\delta(i)$. Thus,

$\tau_{i+1} \leq \tau_i - 2^{-1}L^{-2}\ (1-\lambda)^2\ \delta^2(i) \leq \tau_i - 2^{-1}L^{-2}(1-\lambda)^2\delta^2(j_l),\ i \in I_l$.

Because $0 \leq \tau_i \leq 2D^2$ (evident), the latter inequality immediately implies that the number N_l of elements in I_l satisfies the estimate

$$N_l \leq 4D^2L^2(1-\lambda)^{-2}\delta^{-2}(j_l). \tag{LM.5}$$

B.5. From the definitions of N and of a group, we have

$$\delta(j_1) = \delta(N) > \varepsilon,\ \delta(j_{l+1}) > (1-\lambda)^{-1}\delta(j_l).$$

These relations combined with (LM.5) imply $N = \sum_{l\geq 1} N_l \leq 4D^2L^2(1-\lambda)^{-2} \sum_{l\geq 1} \varepsilon^{-2}\ (1-\lambda)^{2(l-1)} = 4\ (V/\varepsilon)^2\ (1-\lambda)^{-2}\ \lambda^{-1}\ (2-\lambda)^{-1}$. ∎

3.2.2. Dual Level Method (DLM)

A. Description of DLM

Parameters: $\lambda,\ \mu \in (0,1)$

Initialization: (x_1,y_1) is an arbitrary point of $Q\times H$

i-th step:

1) Call the oracle, (x_i,y_i) being the input

2) Compute $f_*(i)$, $\{t_i(j),s_i(j)\}_{1\le j\le i}$

(see 3.2.1.3))

3) Set

$$l(i) = f_*(i) + \lambda\ \Delta(i)\ (= -(1-\lambda)\Delta(i)),$$

$$\varepsilon^+(i) = f_i(x_i,y_i) - l(i) - \mu\ (1-\lambda)\ \Delta(i)$$

(note that $f_i(x_i,y_i) \ge 0$, see (3.4), so that $\varepsilon^+(i) \ge 0$).

Define $p_i \in \mathbb{R}^n\times\mathbb{R}^{n'}$ as the solution to the problem

P(i): $\quad$ *minimize* $|p|^2$ *subject to* $p \in \partial_{\varepsilon^+(i)} f_i(x_i,y_i)$

and set

$$(x_{i+1},y_{i+1}) = \pi((x_i,y_i) - \mu(1-\lambda)\Delta(i)|p_i|^{-2}p_i,Q\times H).$$

The i-th approximate solution is defined as

$$(x_i^* = \sum_{j=1}^{i} s_i(j)\ x_j,\ y_i^* = \sum_{j=1}^{i} t_i(j)\ y_j),$$

where $\{s_i(j)\}_j$ and $\{t_i(j)\}_j$ are the same as in **3.2.1**, namely, the optimal dual solutions to $P_y(i)$, $P_x(i)$, respectively.

B. Efficiency estimate. We claim that

$$\varepsilon(x_i^*,y_i^*) \le -\ f_*(i),$$

$$i \ge c(\lambda,\mu)\ (V/\varepsilon)^2 \Rightarrow \varepsilon(x_i^*,y_i^*) \le \varepsilon,$$

where

$$c(\lambda,\mu) = 4\mu^{-2}(1-\lambda)^{-2}\ \lambda^{-1}\ (2-\lambda)^{-1}$$

(note that $\min_\lambda c(\lambda,\mu) = 16\ \mu^{-2} = c(0.29289...,\mu)$).

Proof.

B.1. The efficiency estimate

$$\varepsilon(x_i^*,y_i^*) \le \Delta(i) \qquad \text{(DLM.1)}$$

was established in 3.2.1.B.1.

B.2. Set $S_i = [f_*(i),0]$. Then (see (3.3), (3.4)) $S_i \neq \varnothing$ and

$$S_1 \supseteq S_2 \supseteq \ldots, \ |S_i| = \Delta(i), \qquad \text{(DLM.2)}$$

where $|S|$ denotes the length of a segment S.

B.3. Let us fix $\varepsilon > 0$ and assume that for certain N and all $i \le N$ we have $\Delta(i) > \varepsilon$. Let us split the integer segment $I = 1,\ldots,N$ in groups $I_1,\ldots, I_k$ as follows. The last element of the first group is $j_1 \equiv N$, and this group contains precisely those $i \in I$ for which $\Delta(i) \le (1-\lambda)^{-1}\Delta(j_1)$. The largest element of I, j_2, which does not belong to the group I_1, if such an element exists, is the last element of I_2, and the latter group consists precisely of those $i \le j_2$, for which $\Delta(i) \le (1-\lambda)^{-1}\ \Delta(j_2)$. The largest element of I, j_3, which does not belong to I_2, is the last element of I_3, and this group consists of those $i \le j_3$ satisfying $\Delta(i) \le (1-\lambda)^{-1}$ $\Delta(j_3)$, and so on.

Let $(u(l),v(l))$ minimize the function $f_{j_l}(\cdot,\cdot)$ over $Q\times H$. For $i \in I_l$ from (DLM.2), the definition of $l(i)$ and the relation $\Delta(j_l) \ge (1-\lambda)\ \Delta(i)$, $i \in I_l$, it immediately follows that $f_*(j_l)) = f_{j_l}(u(l),v(l)) \le l(i)$ for all $i \in I_l$. (3.3) shows that $f_j(u(l),v(l)) \le l(i)$ for all $i,j \in I_l$. Thus, we have established the following:

the (clearly convex) level sets $Q_i = \{z \in Q\times H \mid f_i(z) \le l(i)\}$ associated with $i \in I_l$, have a common point (namely, $z(l) = (u(l),$ $v(l)))$. (DLM.3)

B.4. Let $i \in I_l$, $z_i = (x_i,y_i)$. By virtue of the definition of p_i we have for $z \in Q\times H$: $f^i(z) \equiv f_i(z_i) + p_i^T (z-z_i) - \varepsilon^+(i) \le f_i(z)$. In particular, $f^i(z(l)) \le f_i(z(l)) \le l(i)$, while $f^i(z_i) = f_i(z_i) - \varepsilon^+(i) = l(i) + \mu (1-\lambda) \Delta(i) \ge l(i)$. We conclude that $f^i(z_i) - f^i(z(l)) \ge \mu(1-\lambda)\Delta(i)$, so that $p_i^T(z_i-z(l)) \ge \mu(1-\lambda)\Delta(i)$. Since $z_{i+1} = \pi(z_i - \mu(1-\lambda)\Delta(i) \, |p_i|^{-2} \, p_i, Q\times H)$, it follows that

$$\tau_{i+1} \equiv |z_{i+1} - z(l)|^2 \le \tau_i - |p_i|^{-2} \mu^2(1-\lambda)^2\Delta^2(i);$$

clearly, $|p_i| \le 2^{1/2}L$ (since f_i is Lipschitz continuous with constant $2^{1/2}L$) and $\Delta(i) \ge \Delta(j_l)$, and we obtain

$$\tau_{i+1} \equiv |z_{i+1}-z(l)|^2 \le \tau_i - 2^{-1}L^{-2} \mu^2(1-\lambda)^2\Delta^2(j_l), \; i \in I_l. \tag{DLM.4}$$

Because of $0 \le \tau_i \le 2D^2$ (evident), the latter inequality immediately implies that the number N_l of elements in I_l satisfies the estimate

$$N_l \le 4D^2L^2\mu^{-2}(1-\lambda)^{-2}\Delta^{-2}(j_l). \tag{DLM.5}$$

B.5. From the definitions of N and of a group, we have

$$\Delta(j_1) = \Delta(N) > \varepsilon, \; \Delta(j_{l+1}) > (1-\lambda)^{-1}\Delta(j_l).$$

These relations combined with (DLM.5) imply $N = \sum_{l\ge 1} N_l \le 4D^2 L^2\mu^{-2}(1-\lambda)^{-2} \sum_{l\ge 1} \varepsilon^{-2} (1-\lambda)^{2(l-1)} = 4(V/\varepsilon)^2\mu^{-2}(1-\lambda)^{-2} \lambda^{-1} (2-\lambda)^{-1}$. ■

3.2.3. Truncated Level Method (TLM)

A. Description of TLM

Parameters: $\lambda \in (0,1)$

Initialization: (x_1,y_1) is an arbitrary point of $Q\times H$

***i*-th step:**

1) Call the oracle, (x_i,y_i) being the input

2) Compute $\phi_*(i)$, i.e., solve the convex programming problem

$P_{x,y}(i)$: *minimize*

$\phi_i(x,y) = \max\{(f'_x(x_j,y_j))^T(x-x_j) - (f'_y(x_j,y_j))^T(y-y_j) \mid 1 \le j \le i\}$

subject to $(x,y) \in Q\times H$.

3) Set

$$l(i) = -(1-\lambda)\, \delta(i),$$

and set

$$(x_{i+1},y_{i+1}) = \pi((x_i,y_i), \{(x,y) \mid (x,y) \in Q\times H,\ \phi_i(x,y) \le l(i)\}).$$

The i-th approximate solution is defined as follows:

$$(x^*_i = \sum_{j=1}^{i} r_i(j)\, x_j,\ y^*_i = \sum_{j=1}^{i} r_i(j)\, y_j),$$

where the quantities $\{r_i(j)\}_{1\le j\le i}$ form an optimal dual solution to $P_{x,y}(i)$, i.e., these quantities satisfy the relations

$$\sum_{j=1}^{i} r_i(j) = 1,$$

$$r_i(j) \ge 0,$$

$$\min_{(x,y)\in Q\times H} \sum_{j=1}^{i} r_i(j)\{(f'_x(x_j,y^T_j))(x-x_j) - (f'_y(x_j,y_j))^T (y - y_j) = \min_{Q\times H} \phi_i(\cdot,\cdot).$$

B. Efficiency estimate. We claim that

$$\varepsilon(x^*_i,y^*_i) \le \delta(i),$$

$$i > c(\lambda)\, (V/\varepsilon)^2 \Rightarrow \varepsilon(x^*_i,y^*_i) \le \varepsilon,$$

where

$$c(\lambda) = 4\, (1-\lambda)^{-2}\, \lambda^{-1}\, (2-\lambda)^{-1}$$

(note that $\min c(\cdot) = 16 = c(0.29289...)$).

Proof.

B.1. Let $(x,y) \in Q\times H$. We have

$$f(x_j,y) \le f(x_j,y_j) + (f'_y(x_j,y_j))^T(y-y_j),$$

$$f(x,y_j) \ge f(x_j,y_j) + (f'_x(x_j,y_j))^T(x-x_j),$$

whence

$$f(x,y_j) - f(x_j,y) \geq (f'_x(x_j,y_j))^T(x-x_j) - (f'_y(x_j,y_j))^T(y-y_j),$$

which in turn implies

$$\sum_{j=1}^{i} r_i(j)\ (f(x,y_j) - f(x_j,y)) \geq \sum_{j=1}^{i} r_i(j)\{(f'_x(x_j,y_j))^T(x-x_j) - (f'_y(x_j,y_j))^T(y-y_j)\}.$$

Since f is convex in x and concave in y, we have $f(x,y_i^*) - f(x_i^*,y) \geq \sum_{j=1}^{i} r_i(j)\ (f(x,y_j) - f(x_j,y))$, so that

$$f(x,y_i^*) - f(x_i^*,y) \geq \sum_{j=1}^{i} r_i(j)\{(f'_x(x_j,y_j))^T(x-x_j) - (f'_y(x_j,y_j))^T (y-y_j)\}.$$

Taking the minimum over $(x,y) \in Q\times H$ and using the definition of $r_i(\cdot)$, we obtain

$$\overline{f}(x_i^*) - \underline{f}(y_i^*) \leq \min_{Q\times H}\ \phi_i(\cdot,\cdot).$$

In other words,

$$\varepsilon(x_i^*,y_i^*) \leq \delta(i), \qquad \text{(TLM.1)}$$

as is required in the accuracy estimate.

B.2. Set $S_i = [\phi_*(i),0]$. Then (see (3.6)) $S_i \neq \varnothing$ and

$$S_1 \supseteq S_2 \supseteq \ldots,\ |S_i| = \delta(i), \qquad \text{(TLM.2)}$$

where $|S|$ denotes the length of a segment S.

B.3. Let us fix $\varepsilon > 0$ and assume that for certain N and all $i \leq N$ we have $\delta(i) > \varepsilon$. Let us split the integer segment $I = 1,\ldots,N$ in groups $I_1,\ldots, I_k$ as follows. The last element of the first group is $j_1 \equiv N$, and this group contains precisely those $i \in I$ for which $\delta(i) \leq (1-\lambda)^{-1}\delta(j_1)$. The largest element of I, j_2, which does not belong to the group I_1, if such an element exists, is the last element of I_2, and the latter group consists precisely of those $i \leq j_2$, for which $\delta(i) \leq (1-\lambda)^{-1}\ \delta(j_2)$. The largest element of I, j_3, which does not belong to I_2, is the last element of I_3, and this group consists of those $i \leq j_3$ satisfying $\delta(i) \leq (1-\lambda)^{-1}$

$\delta(j_3)$, and so on.

B.4. From (TLM.2) it immediately follows that $\phi_*(j_l) \le l(i)$, $i \in I_l$. Let $z(l)$ minimize the function $\phi_{j_l}(\cdot)$ over $Q\times H$; then for $i \in I_l$ one has $\phi_i(z(l)) \le \phi_{j_l}(z(l)) \le l(i)$ (see (3.5)). Thus, we have established that

the (clearly convex) level sets $Q_i = \{z \in Q\times H \mid \phi_i(z) \le l(i)\}$ *associated with* $i \in I_l$, *have a common point (namely, the point* $z(l)$*).*

(TLM.3)

B.5. By virtue of the standard properties of the projection mapping, (TLM.3) implies, under the notation $z_i = (x_i, y_i)$,

$$\tau_{i+1} \equiv |z_{i+1}-z(l)|^2 \le \tau_i - \mathrm{dist}^2\{z_i,Q_i\},\ i \in I_l. \qquad \text{(TLM.4)}$$

We also have $\phi_i(z_i) - l(i) \ge -l(i)$ (since clearly $\phi_i(z_i) \ge 0$), so that $\phi_i(z_i) - l(i) \ge (1-\lambda)\delta(i)$, while $\phi_i(z_{i+1}) \le l(i)$. Since ϕ_i is Lipschitz continuous with the constant $2^{1/2}L$ (Remark 3.3), it follows that $\mathrm{dist}\{z_i,Q_i\} = |z_i - z_{i+1}| \ge L^{-1}|\phi_i(z_i) - \phi_i(z_{i+1})| \ge L^{-1}(1-\lambda)\delta(i)$. Thus,

$$\tau_{i+1} \le \tau_i - 2^{-1}L^{-2}(1-\lambda)^2\delta^2(i) \le \tau_i - 2^{-1}L^{-2}(1-\lambda)^2\delta^2(j_l),\ i \in I_l.$$

Because $0 \le \tau_i \le 2D^2$ (evident), the latter inequality immediately implies that the number N_l of elements in I_l satisfies the estimate

$$N_l \le 4D^2L^2(1-\lambda)^{-2}\delta^{-2}(j_l). \qquad \text{(TLM.5)}$$

B.6. From the definitions of N and of a group, we have

$$\delta(j_1) = \delta(N) > \varepsilon,\ \delta(j_{l+1}) > (1-\lambda)^{-1}\delta(j_l).$$

These relations combined with (TLM.5) imply $N = \sum_{l\ge1} N_l \le 4D^2L^2(1-\lambda)^{-2} \sum_{l\ge1} \varepsilon^{-2}(1-\lambda)^{2(l-1)} = 4(V/\varepsilon)^2(1-\lambda)^{-2}\lambda^{-1}(2-\lambda)^{-1}$. ∎

4. Methods for (CMin)

4.0. Additional assumption. In what follows we assume that there exists $x \in Q$ with $G(x) > 0$, so that the problem really is a constrained one.

4.1. Notation. Assume we have called the oracle at the points $x_1,\dots,x_i \in Q$. Then the following objects are defined:

Model of f:

$$f_i(x) = \max\{f(x_j) + (f'(x_j))^T(x-x_j) \mid 1 \le j \le i\}$$

Model of G:

$$G_i(x) = \max\{g_k(x_j) + (g_k'(x_j))^T(x-x_j) \mid 1 \le j \le i,\ 1 \le k \le m\}$$

Remark 4.1.1. Clearly,

$$f_1(x) \le f_2(x) \le \dots \le f_i(x) \le f(x),\ x \in Q \tag{4.1}$$

$$G_1(x) \le G_2(x) \le \dots \le G_i(x) \le G(x),\ x \in Q \tag{4.2}$$

$$f_i(x_j) = f(x_j),\ G_i(x_j) = G(x_j),\ 1 \le j \le i \tag{4.3}$$

and the functions f_i, G_i are Lipschitz continuous with Lipschitz constant L.

Model's best value: $f_*(i) = \min\{f_i(\cdot) \mid x \in Q,\ G_i(x) \le 0\}$

Remark 4.1.2. From Remark 4.1 it follows immediately that $f_*(i)$ are well-defined and

$$f_*(1) \le f_*(2) \le \dots \le f_*(i) \le f^* \tag{4.4}$$

Admissible set: $T(i) = \{(f(x_j),G(x_j)) \mid 1 \le j \le i\} \subset \mathbb{R}^2$

Completed admissible set: $C(i) = (\mathrm{Conv}\ T(i)) + \mathbb{R}^2_+$

4.2. Constrained Level Method (CLM)

4.2.1. Preliminary remarks. Assume we have called the oracle at the points $x_1,\dots,x_i \in Q$. Then, besides the objects described in 4.1, we can define also the following:

Support function:

$$h_i(\alpha) \equiv \min\{\alpha\ (f(x_j)-f_*(i)) + (1-\alpha)\ G(x_j)\ \mid\ 1 \le j \le i\} =$$

$$= \min\{\alpha(u-f_*(i)) + (1-\alpha)\ v\ \mid\ (u,v) \in C(i)\}:\ [0,1] \to \mathbb{R}.$$

Gap:

$$\Delta(i) = \max\{h_i(\alpha)\ \mid\ 0 \le \alpha \le 1\}$$

Best point: let $(u(i),v(i)) \in \operatorname{Argmin}\{\rho(u-f_*(i),v)\ \mid\ (u,v) \in C(i)\}$, where

$$\rho(p,q) = \max\{(p)_+,(q_+)\},$$

Then there clearly exists a convex combination $\sum_{j=1}^{i} r_i(j)\ (f(x_j), G(x_j))$ of points belonging to $T(i)$, such that $\sum_{j=1}^{i} r_i(j)\ (f(x_j), G(x_j)) \le (u(i),v(i))$. Set

$$x_i^* = \sum_{j=1}^{i} r_i(j)\ x_j;$$

this is *the best point* associated with $x_1,\dots,x_i$.

Remark 4.2.1.1.

1). We have

$$x_i^* \in Q,\ \varepsilon(x_i^*) \le \min\{\rho(u-f_*(i),v)\ \mid\ (u,v) \in C(i)\} = \Delta(i). \tag{4.5}$$

□ The inclusion in (4.5) is evident. The inequality follows from the relations $(f(x_i^*)-f_*(i),G(x_i^*)) \le \sum_{j=1}^{i} r_i(j)\ (f(x_j)-f_*(i), G(x_j)) \le (u(i)-f_*(i),v(i)) \le \min\{\rho(u-f_*(i),v)\ \mid\ (u,v) \in C(i)\}\ (1,1)$ (we have taken into account the convexity of f and G). Since $f^* \ge f_*(i)$ (see (4.4)), the resulting inequality implies the inequality in (4.5).

Now let us prove that

$$\Delta(i) = \min\{\rho(u-f_*(i),v)\ \mid\ (u,v) \in C(i)\} \tag{4.6}$$

Indeed, $\rho(p,q) = (\max\{\alpha p + (1-\alpha)q\ \mid\ 0 \le \alpha \le 1\})_+$, whence

$$\min_{(u,v)\in C(i)} \left(\max_{\alpha\in[0,1]} \{\alpha(u-f_*(i))+(1-\alpha)v\} \right)_+ = \left(\min_{(u,v)\in C(i)} \max_{\alpha\in[0,1]} \right.$$

$\{\alpha(u-f_*(i)) + (1-\alpha)v\}\Big)_+ = \left(\max_{\alpha\in[0,1]} \min_{(u,v)\in C(i)} \{\alpha(u-f_*(i))+(1-\alpha)v\}\right)_+$ $= \left(\max_{\alpha\in[0,1]} h_i(\alpha)\right)_+ = (\Delta(i))_+$. It remains to verify that $\Delta(i) \ge 0$. Assume that $\Delta(i) < 0$. Then, evidently, the closed convex set $C(i) \subset \mathbb{R}^2$ cannot be separated (even nonstrictly) from the point $z = (f^*(i),0)$, so that the latter point belongs to the interior of $C(i)$, and, consequently, there exists a convex combination $z' = \sum_{j=1}^{i} t_j\,(f(x_j),G(x_j))$ such that $z' < z$. Since $\sum_{j=1}^{i} t_j\,(f(x_j),G(x_j)) \ge (f_i(\sum_{j=1}^{i} t_j\, x_j),G_i(\sum_{j=1}^{i} t_j\, x_j))$, we obtain $(f_i(\sum_{j=1}^{i} t_j\, x_j),G_i(\sum_{j=1}^{i} t_j\, x_j)) \le (f_*(i),0)$, which contradicts the definition of $f_*(i)$. ■

2). One has

$$h_1(\alpha) \ge h_2(\alpha) \ge \dots\,,\ \alpha \in [0,1], \tag{4.7}$$

and $h_i(\cdot)$ is concave Lipschitz continuous function with Lipschitz constant V.

□ The monotonicity of $h_i(\cdot)$ in i immediately follows from (4.1), (4.2) and (4.4). Since f_i is Lipschitz continuous with constant L and $f(x_j) = f_i(x_j)$, $j \le i$, we have $|f(x_j)-f_*(i)| \le V$, and since G is Lipschitz continuous with the same constant and takes on Q positive (see 4.0) as well as nonpositive (since the problem is consistent) values, we have $|G(x_j)| \le V$, so that $h_i(\cdot)$ is Lipschitz continuous with the constant V. The concavity of h is evident. ■

4.2.2. Description of CLM

Parameters: λ, $\mu \in (0,1)$

Initialization: x_1 is an arbitrary point of Q, $\alpha_{\min}(0) = 0$, $\alpha_{\max}(0) = 1$, $\alpha(1) = 1/2$.

***i*-th step:**

1) Call the oracle, x_i being the input

2) Compute $f_*(i)$, $h_i(\cdot)$, $\Delta(i)$, x_i^*

3) Define $\alpha_{\min}(i)$ as the minimal, and $\alpha_{\max}(i)$ - as the maximal of $\alpha \in [0,1]$ such that $h_i(\alpha) \geq 0$. Set

$$\alpha(i+1) = \begin{cases} (\alpha_{\min}(i)+\alpha_{\max}(i))/2, & \text{if } (\alpha(i)-\alpha_{\min}(i))/(\alpha_{\max}(i) - \alpha_{\min}(i)) \in [\mu/2, 1-\mu/2] \\ \alpha(i), & \text{otherwise} \end{cases}$$

4) set

$$w(i) = \alpha(i)\, f_*(i),\ W(i) = \min_{1\leq j\leq i} (\alpha(i)\, f(x_j) + (1-\alpha(i))\, G(x_j)),$$

$$l(i) = w(i) + \lambda\ (W(i) - w(i)),$$

$$x_{i+1} = \pi(x_i, \{x \mid x \in Q,\ \alpha(i)f_i(x) + (1-\alpha(i))G_i(x) \leq l(i)\}).$$

4.2.3. Efficiency estimate. We claim that

$$\varepsilon(x_i^*) \leq \Delta(i),$$

and if $0 < \varepsilon < V$, then the following implication holds:

$$i > c(\lambda,\mu)\ (V/\varepsilon)^2 \ \ln(2V/\varepsilon) \Rightarrow \varepsilon(x_i^*) \leq \varepsilon,$$

where

$$c(\lambda,\mu) = 2\ (\ln 2)^{-1}\ (1+1/\mu)^2\{\ln(2/(1+\mu))^{-1}(1-\lambda)^{-2}(2-\lambda)^{-1}\lambda^{-1}$$

(note that $\min c(\cdot,\cdot) = c(0.29289...,0.53247...) \leq 360$).

Proof.

1) The efficiency estimate

$$\varepsilon(x_i^*) \leq \Delta(i)$$

was established in Remark 4.2.1.1.1.

2) Let $\varepsilon > 0$ and let N be such that $\Delta(i) > \varepsilon$. Let us split the integer segment $I = 1,...,N$ into sequential groups $J_1,...,J_k$ in such a way that $\alpha(i) \equiv \alpha_l$ is constant for $i \in J_l$ and $\alpha_l \neq \alpha_{l+1}$. Let p_l be the first, and q_l be the last element of J_l. We call a group *substantial*, if $q_l > p_l$.

3) Let us prove that the amount k of groups satisfies the

relation

$$k \le \{\ln(2/(1+\mu))\}^{-1} \ln(V/\varepsilon + 1) + 1. \qquad \text{(CLM.1)}$$

Indeed, let $T_0 = [0,1]$, $T_i = [\alpha_{\min}(i), \alpha_{\max}(i)]$, $i \ge 1$. Then $T_i \supseteq T_{i+1}$ (see (4.7)) and $h_i(\cdot)$ is negative outside T_i. Note that α_l is the center of T_{p_l-1} and for $l < k$, either α_l does not belong to T_{q_l}, or this segment is divided by α_l into parts such that at least one of them is less than $\mu|T_{q_l}|/2$. Since $T_{q_l} \subset T_{p_l-1}$, it follows that $|T_{q_l}| \le (1+\mu)|T_{p_l-1}|/2 = (1+\mu)|T_{q_{l-1}}|/2$, where $q_0 = 0$. Thus, if $k > 1$, then $|T_N| \le |T_{q_{k-1}}| \le ((1+\mu)/2)^{k-1}$. Since $h_N(\cdot)$ is negative outside T_N and is Lipschitz continuous with the constant V (Remark 4.2.1.1.2)), it follows that in the case of $k > 1$ we have $\Delta(N) = \max\limits_{0\le\alpha\le 1} h_N(\alpha) \le V\,((1+\mu)/2)^{k-1}$. Since $\Delta(N) > \varepsilon$, we obtain in the case of $k > 1$: $k \le \{\ln(2/(1+\mu))\}^{-1} \ln(V/\varepsilon) + 1$, which implies (CLM.1).

4) Now let us prove that the amount of elements, M_l, in the group J_l satisfies the relation

$$M_l \le 1 + (1+1/\mu)^2(1-\lambda)^{-2}(2-\lambda)^{-1}\lambda^{-1}\,(V/\varepsilon)^2. \qquad \text{(CLM.2)}$$

Of course, we can assume that the group J_l under consideration is substantial. Denote $J_l' = J_l \setminus \{q_l\}$. Let $\delta(i) = h_i(\alpha_l)$ ($= W(i) - w(i)$). We have (see (4.4) and (4.7))

$$W(p_l) \ge W(p_l+1) \ge \dots \ge W(q_l-1), \qquad \text{(CLM.3)}$$

$$w(p_l) \le \dots \le w(q_l-1), \qquad \text{(CLM.4)}$$

so that

$$\delta(p_l) \ge \delta(p_l+1) \ge \dots \ge \delta(q_l-1). \qquad \text{(CLM.5)}$$

Let us prove that

$$\delta(q_l-1) \ge (\mu/(\mu+1))\,\varepsilon. \qquad \text{(CLM.6)}$$

Indeed, α_l splits the segment T_{q_l-1} in two parts, each not shorter than $\mu|T_{q_l-1}|/2$; $h_{q_l-1}(\cdot)$ is nonnegative on T_{q_l-1} and concave, so that $\max_{T_{q_l-1}} h_{q_l-1}(\cdot) \le (1 + 1/\mu)\ h_{q_l-1}(\alpha_l) = (1 + 1/\mu)\ \delta(q_l-1)$. Outside T_{q_l-1} the function $h_{q_l-1}(\cdot)$ is negative; thus, $\Delta(q_l-1) = \max_{T_{q_l-1}} h_{q_l-1}(\cdot) \le (1 + 1/\mu)\ \delta(q_l-1)$. Since $\Delta(q_l-1) > \varepsilon$, we obtain (CLM.6).

5) Let us split the integer segment J'_l into groups $I_1,\dots,I_s$ as follows. The last element of I_1 is $j_1 \equiv q_l - 1$, and I_1 consists precisely of those $i \in J'_l$ for which $\delta(i) \le (1-\lambda)^{-1}\ \delta(j_1)$. The largest $i \in J'_l$ which does not belong to I_1, if such an i exists, is the last element, j_2, of the second subgroup I_2, and I_2 consists precisely of those $i \in J'_l$, $i \le j_2$, for which $\delta(i) \le (1-\lambda)^{-1}\ \delta(j_2)$. The last element of J'_l which does not belong to $I_1 \cup I_2$, if such an element exists, is the last element, j_3, of I_3, and I_3 consists of those $i \in I\backslash(I_1 \cup I_2)$ for which $\delta(i) \le (1 - \lambda)^{-1}\ \delta(j_3)$, and so on. Let us prove that the number of elements, N_r, in the subgroup I_r, satisfies the relation

$$N_r \le D^2L^2\ (1-\lambda)^{-2}\ \delta^{-2}(j_r). \qquad \text{(CLM.7)}$$

Indeed, let $i \in I_r$ and let $S_i = [w(i),W(i)]$. Then (see (CLM.3) - (CLM.6)) S_i are nonempty segments, $|S_i| = \delta(i)$; besides this, $S_{i+1} \subseteq S_i$, $i+1 \in I_l$. Let $\phi^l(x) = \alpha_l f(x) + (1-\alpha_l)\ G(x)$, and let $\phi_i(x) = \alpha_l f_i(x) + (1-\alpha_l)G_i(x)$. Then clearly

$$\phi_{i_r}(\cdot) \le \phi_{i_r+1}(\cdot) \le \dots \le \phi_{j_r}(\cdot) \le \phi^l(\cdot), \qquad \text{(CLM.8)}$$

where i_r is the first element of I_r, and

$$\phi_i(x_i) \ge W(i),\ i \in I_r,\ \min_Q \phi_i(\cdot) \le w(i). \qquad \text{(CLM.9)}$$

Let $u(r)$ minimize $\phi_{j_r}(\cdot)$ over Q. Then (see (CLM.9) $\phi_{j_r}(u(r)) \le w(j_r)$, so that (see (CLM.8)) $\phi_i(u(r)) \le w(j_r)$. On the other hand, for $i \in I_r$ we have $l(i) = w(i) + \lambda\,(W(i) - w(i)) = W(i) - (1-\lambda)\delta(i) \ge W(j_r) - \delta(j_r) = w(j_r)$ (we have taken into account that $W(i) \ge W(j_r)$ and $\delta(i) \le (1-\lambda)^{-1}\,\delta(j_r)$, $i \in I_r$). Thus, $\phi_i(u(r)) \le l(i)$, $i \in I_r$. We have proved that

the (clearly convex) level sets $Q_i = \{x \in Q \mid \phi_i(x) \le l(i)\}$, $i \in I_r$, *have a common point (namely,* $u(r)$*).* (CLM.10)

Now, $x_{i+1} = \pi(x_i,Q_i)$, $i \in I_r$. In view of the standard properties of the projection mapping, we have

$$\tau_{i+1} \equiv |x_{i+1} - u(r)|^2 \le \tau_i - \mathrm{dist}^2(x_i,Q_i). \qquad \text{(CLM.11)}$$

Furthermore, $\phi_i(x_i) \ge W(i)$ (see (CLM.9)) and $\phi_i(x_{i+1}) \le l(i)$, so that $\phi_i(x_i) - \phi_i(x_{i+1}) \ge (1-\lambda)\;\delta(i)$. Clearly, $\phi_i(\cdot)$ is Lipschitz continuous with the constant L, and we obtain that $|x_i - x_{i+1}| = \mathrm{dist}(x_i,Q_i) \ge L^{-1}\,(1-\lambda)\;\delta(i)$. Thus, (CLM.11) implies

$$\tau_{i+1} \le \tau_i - L^{-2}(1-\lambda)^2\delta^2(i) \le \tau_i - L^{-2}(1-\lambda)^2\delta^2(j_r),\ i \in I_r.$$

Since clearly $\tau_\cdot \le D^2$, (CLM.7) follows.

It remains to note that $\delta(j_{r+1}) > (1-\lambda)^{-1}\;\delta(j_r)$, so that $M_l = |J_l'| + 1 = 1 + \sum_r N_r \le 1 + D^2L^2\;\delta^{-2}(j_1)\;(1-\lambda)^{-2}(2-\lambda)^{-1}\lambda^{-1}$, which combined with (CLM.6) proves (CLM.2).

6) (CLM.2) combined with (CLM.1) imply the required efficiency estimate. ∎

4.3. Constrained Newton Method (CNM)

4.3.1. Preliminary remarks. Denote

$$F_t(x) = \rho(f(x)-t,G(x)),$$

where, as above, $\rho(u,v) = \max\{(u)_+,(v)_+\}$, and let

$$\kappa(t) = \min_Q F_t(\cdot).$$

Assume we have called the oracle at the points $x_1,\dots,x_i \in Q$. Then, besides the objects described in 4.1, we can define also the following:

Upper distance function:

$$\kappa^*(i;t) \equiv \min\{\rho(u-t,v) \mid (u,v) \in C(i)\}$$

Lower distance function:

$$\kappa_*(i;t) \equiv \min\{\rho(f_i(x)-t,G_i(x)) \mid x \in Q\}$$

Remark 4.3.1.1. The functions $\kappa(t)$, $\kappa^*(i;t)$, $\kappa_*(i;t)$ are nonincreasing convex Lipschitz continuous with Lipschitz constant 1 functions of $t \in \mathbb{R}$, and

$$\kappa_*(1;t) \le \kappa_*(2;t) \le \dots \le \kappa_*(i;t) \le \kappa(t), \tag{4.8}$$

$$\kappa^*(1;t) \ge \kappa^*(2;t) \ge \dots \ge \kappa^*(i;t) \ge \kappa(t). \tag{4.9}$$

□ $\rho(\cdot,\cdot)$ is monotone and convex on $\mathbb{R}^2$; therefore for convex $p(\cdot)$, $q(\cdot)$: $Q \to \mathbb{R}$ the function $\rho(p(x)-t,q(x))$ is convex on $Q\times\mathbb{R}$, so that $\min_Q \rho(p(\cdot)-t,q(\cdot))$ is convex on $\mathbb{R}$ (and clearly Lipschitz continuous with constant 1). These remarks prove the convexity and the Lipschitz continuity of κ, κ^*, κ_*. The monotonicity of κ^* and κ_* in i immediately follow the monotonicity of ρ combined with (4.1), (4.2) and the (evident) inclusions $C(1) \subset C(2) \subset \dots \subset C(i)$. (4.1), (4.2) and the monotonicity of ρ imply also the inequality $\kappa_*(i;t) \le \kappa(t)$. Convexity of f and G implies immediately that for every $(u,v) \in C(i)$ there exists a convex combination x of the points $x_1,\dots,x_i$ such that $(f(x),G(x)) \le (u,v)$, and this observation combined with the monotonicity of ρ, leads to the inequality $\kappa^*(i;t) \ge \kappa(t)$. ■

Best point: let $(u_i(t),v_i(t)) \in \text{Argmin}\{\rho(u-t,v) \mid (u,v) \in$

$C(i)\}$. Then there clearly exists a convex combination $\sum_{j=1}^{i} r_i(j;t)\ (f(x_j),G(x_j))$ of points belonging to $T(i)$, such that $\sum_{j=1}^{i} r_i(j;t)\ (f(x_j),G(x_j)) \le (u_i(t),v_i(t))$. Set

$$x_i^*(t) = \sum_{j=1}^{i} r_i(j;t)\ x_j;$$

Remark 4.3.1.2. Let $t \le f^*$. Then

$$\varepsilon(x_i^*(t)) \le \kappa^*(i;t). \tag{4.10}$$

□ Indeed, we have $(f(x_i^*(t))-t,G(x_i^*(t))) \le \sum_{j=1}^{i} r_i(j;t)\ (f(x_j)-t,G(x_j)) \le (u_i(t)-t,v_i(t))$, so that $\rho(f(x_i^*(t))-t,G(x_i^*(t))) \le \rho(u_i(t)-t,v_i(t)) = \kappa^*(i;t)$. It remains to note that $t \le f^*$, so that $\varepsilon(x_i^*(t)) = \rho(f(x_i^*(t))-f^*,G(x_i^*(t))) \le \rho(f(x_i^*(t))-t,G(x_i^*(t))) \le \kappa^*(i;t)$. ■

4.3.2. Description of CNM

Parameters: $\lambda \in (0,1)$, $\mu \in (1/2,1)$

Initialization: x_1 is an arbitrary point of Q

***i*-th step:**

1) Call the oracle, x_i being the input

2) Compute $f_*(i)$, $\kappa^*(i;\cdot)$, $\kappa_*(i;\cdot)$

3) Set

$$t_i = \begin{cases} f_*(i), & i = 1 \text{ or if } (\kappa_*(i;t_{i-1}) > \mu\ \kappa\ (i;t^*_{i-1})) \\ t_{i-1}, & \text{otherwise.} \end{cases}$$

$$w(i) = \kappa_*(i;t_i),\ W(i) = \kappa^*(i;t_i),$$

$$l(i) = w(i) + \lambda\ (W(i) - w(i)),$$

$$x_{i+1} = \pi(x_i,\{x \mid x \in Q,\ \rho(f_i(x)-t_i,G_i(x)) \le l(i)\}).$$

4.3.3. Efficiency estimate. We claim that

$$\varepsilon(x_i^*(t_i)) \le \kappa^*(i;t_i),$$

and if $0 < \varepsilon < V$, then the following implication holds:

$$i > c(\lambda,\mu)\ (V/\varepsilon)^2\ \ln(18\ V/\varepsilon) \Rightarrow \varepsilon(x_i^*(t_i)) \le \varepsilon,$$

where

$$c(\lambda,\mu) = 2\ \{\ln(2\mu)\}^{-1}\ (1-\mu)^{-2}(1-\lambda)^{-2}\ (2-\lambda)^{-1}\ \lambda^{-1}$$

(note that $\min c(\cdot,\cdot) = c(0.29289...,0.65252...) \le 249$).

Proof.

1) The accuracy estimate follows from Remark 4.3.1.2 combined with the fact that $t_i = f_*(i')$ for each $i \ge 1$ and some i' depending on i (see the description of the method), while $f_*(i) \le f^*$ in view of (4.4).

2) Let $\varepsilon > 0$, and let N be such that $\kappa^*(N;t_N) > \varepsilon$. Let us split the integer segment $I = 1,...,N$ into groups $J_1,...,J_k$ as follows. The first element of J_1 is $p_1 = 1$, and J_1 consists of those $i \in I$ for which $t_i = t_1$. In the case of $I\backslash I_1 \ne \varnothing$ the first element, p_2, of the latter set is the first element of J_2, and J_2 consists of those $i \in I\backslash I_1$ for which $t_i = t_{p_2}$. If $I\backslash(I_1 \cup I_2) \ne \varnothing$, then the first element, p_3, of the latter set is the first element of J_3, and J_3 consists of those $i \in I\backslash(I_1 \cup I_2)$ for which $t_i = t_{p_3}$, and so on.

3) Let us prove that the amount k of the groups $J_1,...,J_k$ satisfies the relation

$$k \le 2 + \{\ln(2\mu)\}^{-1}\ \ln(2\mu V/\varepsilon + 1). \qquad \text{(CNM.1)}$$

Indeed, $t_i = t(l)$, $i \in J_l$. We have (see the description of the method)

$$\left.\begin{array}{l} t(l) = f_*(p_l),\ 1 \le l \le k, \\ \kappa_*(p_l;t(l-1)) > \mu\ \kappa^*(p_l;t(l-1)),\ 1 < l \le k. \end{array}\right\} \qquad \text{(CNM.2)}$$

Note that since $\kappa_*(i;\cdot) \le \kappa^*(i;\cdot)$ (see (4.8) – (4.9)) (CNM.2) implies

$$\kappa_*(p_l;t(l-1)) > 0,\ 1 < l \le k. \tag{CNM.3}$$

Let us prove that

$$t(1) \le t(2) \le \dots \le t(k) \le f^*. \tag{CNM.4}$$

Indeed, the relations $t(i) \le f^*$ were already established (see 1)). Let us prove that $t_i \ge t_{i-1}$, $1 < i \le N$. We have either $t_i = t_{i-1}$, or $\kappa_*(i;t_{i-1}) > \mu\ \kappa^*(i;t_{i-1})$ and $t_i = f_*(i)$. In the latter case, since $\kappa^*(i;t) \ge \kappa_*(i;t)$, we have $\kappa^*(i;t_{i-1}) > 0$ and therefore $\kappa_*(i;t_{i-1}) > 0$. At the same time, by the definition of $f_*(\cdot)$, for every i there exists a depending on i $x^+ \in Q$ such that $f_i(x^+) = f_*(i)$, $G_i(x^+) \le 0$, which combined with the definition of $\kappa_*(i;\cdot)$ implies

$$\kappa_*(i;f_*(i)) \le 0. \tag{CNM.5}$$

Thus, the relations $t_i = f_*(i)$ and $\kappa_*(i;t_{i-1}) > 0$ combined with the fact that $\kappa_*(i;\cdot)$ is a nonincreasing function, imply $t_{i-1} < t_i$.

Let

$$\kappa_l(t) = \kappa_*(p_l;t),\ 1 \le l \le k,$$

$$\delta(l) = -\ \kappa_l(t(l-1))\ \kappa_l'(t(l-1)),\ 1 < l \le k.$$

Since $\kappa_*(i;\cdot)$ is a nonnegative nonincreasing function, we have $\delta(l) \ge 0$.

Let us prove that

$$\kappa_l(t(l-1)) + \kappa_l'(t(l-1))(t(l) - t(l-1)) \le 0,\ 1 < l \le k. \tag{CNM.6}$$

Indeed, assume that $\kappa_l(t(l-1)) + \kappa_l'(t(l-1))(t(l) - t(l-1)) > 0$. Since $\kappa_l(\cdot)$ is convex, it follows that $\kappa_l(t(l)) \equiv \kappa_*(p_l;t(l)) > 0$, or, which is the same in view of (CNM.2), $\kappa_*(p_l;f_*(p_l)) > 0$; the latter relation contradicts (CNM.5).

Since $\kappa_*(i;\cdot)$ is a convex nonincreasing function, we have for $k \geq l > 2$:

$$\kappa_l(t(l-2)) \geq \kappa_l(t(l-1)) + |\kappa'_l(t(l-1))|(t(l-1)-t(l-2)). \tag{CNM.7}$$

We have $\kappa'_{l-1}(t(l-2)) \neq 0$, since otherwise (CNM.6) would imply $\kappa_{l-1}(t(l-2)) \leq 0$, which contradicts (CNM.3). Thus, (CNM.6) implies for $k \geq l > 2$: $t(l-1) - t(l-2) \geq |\kappa'_{l-1}(t(l-2))|^{-1} \kappa_{l-1}(t(l-2))$, or, in view of (CNM.7), $\kappa_l(t(l-2)) \geq \kappa_l(t(l-1)) + |\kappa'_l(t(l-1))|$ $\kappa_{l-1}(t(l-2))$ $|\kappa'_{l-1}(t(l-2))|^{-1}$. Since $\kappa_{l-1}(t(l-2)) > 0$ (see (CNM.3)), we obtain

$$\kappa_l(t(l-2))/\kappa_{l-1}(t(l-2)) \geq$$

$$\geq \kappa_l(t(l-1))/\kappa_{l-1}(t(l-2)) + |\kappa'_l(t(l-1))|/|\kappa'_{l-1}(t(l-2))| \tag{CNM.8}$$

Since $\kappa_l(t(l-2)) \leq \kappa(t(l-2))$, $\kappa_{l-1}(t(l-2)) > \mu\ \kappa(t(l-2))$ (see (4.8), (4.9) and (CNM.2)), we obtain $\kappa_l(t(l-2))/\kappa_{l-1}(t(l-2)) \geq \mu^{-1}$, while the right hand side of (CNM.8) is not less than $2\{\kappa_l(t(l-1))\ |\kappa'_l(t(l-1))|\}^{1/2}/\{\kappa_{l-1}(t(l-2))\ |\kappa'_{l-1}(t(l-2))|\}^{1/2}$. Thus, (CNM.8) implies

$$\delta(l) \leq (2\mu)^{-2}\delta(l-1),\ 2 < l \leq k. \tag{CNM.9}$$

Now we can complete the proof of (CNM.1). Let $k > 2$. We clearly have $\kappa(t_1) \leq V$ and $|\kappa'_l(t)| \leq 1$; therefore from (CNM.9) it follows that $\delta(k) \leq (2\mu)^{-2(k-2)}V$, so that either $\kappa_k(t(k-1)) \leq (2\mu)^{-k+2}\ V$ or $|\kappa'_k(t(k-1))| \leq (2\mu)^{-k+2}$. Since $\kappa_k(f_*(p_k)) \leq 0$ (see (CNM.5)) and κ_k is a concave nonincreasing function, in the second case we have $\kappa_k(t(k-1)) \leq (2\mu)^{-k+2}\ |f_*(p_k)-t(k-1)| \leq (2\mu)^{-k+2}\ V$ (the latter inequality is evident). Thus, in both cases we have

$$\kappa_k(t(k-1)) \leq (2\mu)^{-k+2}\ V. \tag{CNM.10}$$

In view of (CNM.2) the latter relation means that

$$\kappa^*(p_k;t(k-1)) \leq \mu^{-1}\ (2\mu)^{-k+2}\ V,$$

and since $p_k \le N$ and $t(k-1) \le t_N$ (see (CNM.4)), we conclude from (4.8) and the monotonicity of $\kappa^*(N;\cdot)$ that $\kappa^*(N;t_N) \le \mu^{-1} (2\mu)^{-k+2} V$. Thus, $\mu^{-1} (2\mu)^{-k+2} V > \varepsilon$ (definition of N), so that in the case $k > 2$ (CNM.1) does hold. Of course, it also holds in the case $k \le 2$.

4) Now let us prove that the number N_l of elements in the group J_l satisfies the relation

$$N_l \le 1 + (1-\mu)^{-2}(1-\lambda)^{-2} (2-\lambda)^{-1} \lambda^{-1} (V/\varepsilon)^2 \qquad \text{(CNM.11)}$$

Let $J_l = \{p_l, p_l+1, \ldots, q_l\}$. (CNM.11) is evident in the case $q_l = p_l$. In the opposite case let $J'_l = J_l \setminus \{q_l\}$. Observe that, inside J'_l, the method is basically the standard Level method with parameter λ, applied to the function (convex and Lipschitz continuous with constant L)

$$d(x) = \max\{(f(x)-t(l))_+, g_1(x), \ldots, g_m(x)\}: Q \to \mathbb{R},$$

the quantities $w(i)$ being the best model's values. More precisely, the only differences with LM are:

a) more detailed models of $d(\cdot)$: first, we use the known max-structure of the function and take its model as the maximum of the standard models of the maximands $f(x)-t(l), g_1(x), \ldots, g_m(x)$; second, we append to these more detailed models the information obtained at the iterations preceding those from the group J'_l under consideration;

b) instead of best function's values we use some other quantities (namely, $W(i)$), which, first, are not less than the best model's values $w(i)$, second, do not increase with i and, third, satisfy the relations $d(x_i) \ge W(i)$.

From the above theoretical analysis of the basic Level method it follows that these modifications do not influence the efficiency estimate: the number of iterations (in the group J'_l) required to ensure the relation $W(i) - w(i) \le \nu$ does not exceed the quantity $(1-\lambda)^{-2} (2-\lambda)^{-1} \lambda^{-1} (V/\nu)^2 + 1$.

Now note that if j is the last element of J'_l, then $W(j) - w(j) \equiv \kappa^*(j;t_{j-1}) - \kappa_*(j;t_{j-1}) > (1-\mu) \kappa^*(j;t_{j-1})$ (otherwise the group J_l would terminate immediately after the j-th iteration). We also have $\kappa^*(j;t_{j-1}) \ge \kappa^*(N,t_{j-1})$ (see (4.8)) and $\kappa^*(N,t_{j-1}) \ge \kappa^*(N,t_N) > \varepsilon$. Thus, $W(i) - w(i) > (1-\mu) \varepsilon$, so that $i - p_l + 1 \le (1-\mu)^{-2} (1-\lambda)^{-2} (2-\lambda)^{-1} \lambda^{-1} (V/\varepsilon)^2$. It immediately implies (CNM.11).

5) (CNM.1) combined with (CNM.11) implies the required efficiency estimate. ■

5. A Method for (Var)

5.1. Notation. Assume we have called the oracle at the points $x_1,\dots,x_i \in Q$. Then the following objects are defined:

Model:

$$\phi_i(x) = \max\{(F(x_j))^T(x-x_j) \mid 1 \le j \le i\}$$

Model's best value:

$$\phi_*(i) = \min_Q \phi_i(x)$$

Gap:

$$\delta(i) = - \phi_*(i).$$

Optimal multipliers are the quantities $r_i(j)$, $1 \le j \le i$, such that $r_i(j) \ge 0$, $\sum_{j=1}^{i} r_i(j) = 1$, and

$$\min\{ \sum_{j=1}^{i} r_i(j) \, (F(x_j))^T(x-x_j) \mid x \in Q\} = \min_Q \phi_i(\cdot) = \phi_*(i). \quad (5.1)$$

Best point:

$$x_i^* = \sum_{j=1}^{i} r_i(j)\, x_j.$$

Remark 5.1.1.

1). We evidently have

$$\phi_1(x) \le \phi_2(x) \le .. \tag{5.2}$$

and $\phi_i(\cdot)$ are convex and Lipschitz continuous with constant L.

2). We have

$$\delta(1) \ge \delta(2) \ge \ldots \ge 0 \tag{5.3}$$

□ Indeed, let x^* be a solution to (Var), so that $(F(x))^T(x - x^*) \ge 0$, $x \in Q$, whence $\phi_i(x^*) \le 0$ and therefore $\phi_*(i) = -\delta(i) \ge 0$. Thus, $\delta(\cdot)$ is positive. The monotonicity of $\delta(i)$ in i follows from (5.2). ■

3). We have

$$\varepsilon(x_i^*) \le \delta(i). \tag{5.4}$$

□ Indeed, let $x \in Q$. Then $(F(x))^T(x-x_i^*) = (F(x))^T \sum_{j=1}^{i} r_i(j)\,(x - x_i^*) \ge \sum_{j=1}^{i} r_i(j)\,(F(x_j))^T(x - x_i^*) \ge \min\{\sum_{j=1}^{i} r_i(j)\,(F(x_j))^T(y - x_i^*) \mid y \in Q\} = \phi_*(i)$ (we have taken into account the monotonicity of $F(\cdot)$ and (5.1)). Thus, $\varepsilon(x_i^*) = \max\{(F(x))^T(x_i^*-x) \mid x \in Q\} \le -\phi_*(i) = \delta(i)$. ■

5.2. Truncated Level Method (TLM) for (Var)

A. Description of TLM

Parameters: $\lambda \in (0,1)$

Initialization: x_1 is an arbitrary point of Q

***i*-th step:**

1) Call oracle, x_i being the input

2) Compute $\phi_*(i)$ and x_i^*

3) Set

$$l(i) = -(1-\lambda)\,\delta(i),$$

$$x_{i+1} = \pi(x_i, \{x \mid x \in Q,\ \phi_i(x) \le l(i)\})$$

B. Efficiency estimate. We claim that

$$\varepsilon(x_i^*) \le \delta(i),$$
$$i > c(\lambda)\ (V/\varepsilon)^2 \Rightarrow \varepsilon(x_i^*) \le \varepsilon,$$

where

$$c(\lambda) = (1-\lambda)^{-2}\ \lambda^{-1}\ (2-\lambda)^{-1}$$

(note that min $c(\cdot) = 4 = c(0.29289)$).

Proof.

B.1. The efficiency estimate

$$\varepsilon(x_i^*) \le \delta(i) \qquad \text{(TLM.1)}$$

was established in (5.4).

B.2. Set $S_i = [\phi_*(i), 0]$. Then (see (5.2), (5.3)) $S_i \ne \emptyset$ and

$$S_1 \supseteq S_2 \supseteq \ldots,\ |S_i| = \delta(i), \qquad \text{(TLM.2)}$$

where $|S|$ denotes the length of a segment S.

B.3. Let us fix $\varepsilon > 0$ and assume that for certain N and all $i \le N$ we have $\delta(i) > \varepsilon$. Let us split the integer segment $I = 1,\ldots,N$ in groups $I_1,\ldots, I_k$ as follows. The last element of the first group is $j_1 \equiv N$, and this group contains precisely those $i \in I$ for which $\delta(i) \le (1-\lambda)^{-1}\delta(j_1)$. The largest element of I, j_2, which does not belong to the group I_1, if such an element exists, is the last element of I_2, and the latter group consists precisely of those $i \le j_2$, for which $\delta(i) \le (1-\lambda)^{-1}\ \delta(j_2)$. The largest element of I, j_3, which does not belong to I_2, is the last element of I_3, and this group consists of those $i \le j_3$ satisfying $\delta(i) \le (1-\lambda)^{-1}$ $\delta(j_3)$, and so on.

B.4. From (TLM.2) it immediately follows that $\phi_*(j_l) \le l(i)$, $i \in I_l$. Let $u(l)$ minimize the function $\phi_{j_l}(\cdot)$ over Q; then for $i \in$

I_l one has $\phi_i(u(l)) \le \phi_{j_l}(u(l)) \le l(i)$. Thus, we have established that

the (clearly convex) level sets $Q_i = \{x \in Q \mid \phi_i(x) \le l(i)\}$ *associated with* $i \in I_l$, *have a common point (namely,* $u(l)$*).* (TLM.3)

B.5. By virtue of the standard properties of the projection mapping, (TLM.3) implies

$$\tau_{i+1} \equiv |x_{i+1}-u(l)|^2 \le \tau_i - \mathrm{dist}^2(x_i,Q_i), \; i \in I_l. \qquad \text{(TLM.4)}$$

We also have $\phi_i(x_i) - l(i) \ge - l(i)$ (see (2.9)), so that $\phi_i(x_l) - l(i) \ge (1-\lambda)\delta(i)$, and $\phi_i(x_{i+1}) \le l(i)$. From the Lipschitz property of ϕ_i, it follows that $\mathrm{dist}(x_i,Q_i) = |x_i - x_{i+1}| \ge L^{-1}|\phi_i(x_i) - \phi_i(x_{i+1})| \ge L^{-1}(1-\lambda)\delta(i)$. Thus,

$$\tau_{i+1} \le \tau_i - L^{-2}(1-\lambda)^2\,\delta^2(i) \le \tau_i - L^{-2}(1-\lambda)^2\delta^2(j_l), \; i \in I_l.$$

Because $0 \le \tau_i \le D^2$ (evident), the latter inequality immediately implies that the number N_l of elements in I_l satisfies the estimate

$$N_l \le D^2L^2(1-\lambda)^{-2}\delta^{-2}(j_l). \qquad \text{(TLM.5)}$$

B.6. From the definitions of N and of a group, we have

$$\delta(j_1) = \delta(N) > \varepsilon, \; \delta(j_{l+1}) > (1-\lambda)^{-1}\delta(j_l).$$

These relations combined with (TLM.5) imply $N = \sum_{l\ge 1} N_l \le D^2L^2(1-\lambda)^{-2} \sum_{l\ge 1} \varepsilon^{-2}(1-\lambda)^{2(l-1)} = (V/\varepsilon)^2(1-\lambda)^{-2}\lambda^{-1}(2-\lambda)^{-1}$. ∎

6. Computational results

All the test-problems described below are available from the authors.

6.1. Unconstrained minimization

We have tested the simplest method of those described in Sect. 2, namely the Level method LM.

Our implementation used two features:

* An input parameter $f_*(0)$ was given to the algorithm, serving as a lower bound on the optimal value f_*. The algorithm could then be run without compactness assumption on Q.

* The two auxiliary problems to compute $f_*(i)$ and x_{i+1} were solved with the help of the code QL0001 of K. Schittkowsky, itself based on the algorithm of [Pow. 1983]. In some of the experiments we used simplex codes of E. Borisova and N. Sokolov in order to compute $f_*(i)$.

In all our experiments reported below, the parameter λ was set to 0.5 and the algorithm was run until the gap became smaller than 10^{-6} (in relative accuracy). We used double precision Fortran on a Sun Workstation. The test-problems were the following:

* **BADGUY**. This is a hand-made function, illustrating worst-case behaviours; see [NYu 1983]. It is organized so that the gap after $i\ n$ calls to the oracle (n is the dimension of the problem) cannot be reduced by more than the factor 2^{3i+1}. We used $n=30$ variables.

* **MAXQUAD** and **TR48** are described in [LM 1978].

* **MAXANAL** is a regularization of MAXQUAD, where the objective $\max\{f_k(x)\}$ is replaced by

$$\max\{\Sigma\lambda_k f_k(x) + \varepsilon\ \Sigma\ \ln(\lambda_k)\ \mid\ \Sigma\lambda_k = 1\}.$$

Here, $\varepsilon=10^{-3}$.

* **NET22h** is the dual of a network problem, described by Goffin. It has 22 variables and is badly scaled.

* **URY100** is a convex variant of a problem defined by Uryasjev. It is actually the sum of a piecewise linear function and of a quadratic, with $n=100$ variables bound by the box $-0.2 \le x_i \le 0.2$.

* **TSP** is the dual of a traveling salesman problem, following the Lagrangian relaxation of [HK 1971]. The function to be minimized is therefore the maximum of a very large number of affine functions; we used datasets with n = *6, 14, 29, 100, 120* and *442* variables respectively, coming from VLSI design.

The results are reported in Tables 1 to 5 (see Appendix 2). Observe the quality of the performances, as compared to the simplicity of the implementation. Generally speaking, the method is comparable to the best known methods, except on TSP442 (where it can be considered as non-convergent). Indeed, a weak point of the approach is to use the (bad) cutting plane model to provide the estimate $f_*(i)$. We have experimented the variant of Level in which $f_*(i)$ is fixed to the optimal value f^* (assumed known). When applied to TSP442, this variant does reach the value -50505.5 (in 500 iterations, and the algorithm was stopped there). This seems to confirm the important role of $f_*(i)$; research is currently in progress for a proper management of it.

6.2. Saddle points

We tested the Level method on a number of randomly generated saddle point problems of the following type:

find a saddle point of the quadratic function

$$f(x,y) = \frac{1}{2}(Px,x) - \frac{1}{2}(Qy,y) + (Rx,y)$$

under the constraints

$$Ax \le a, \ \|x\|_\infty \le r, \ By \le b, \ \|y\|_\infty \le r,$$

where x and y are both n-dimensional, P, Q and R are matrices of corresponding sizes, and P and Q are positive semidefinite. The numbers of rows in the constraint matrices A, B, are equal to m.

We used a simple generator of test problems. The input to the generator includes the sizes n, m as well as the parameter dc used to control the condition numbers of P and Q and the range of Lagrange coefficients at the saddle point (i.e., the coefficients in the representation of f'_x, f'_y at the solution as linear combinations of the gradients of the linear constraints active at the solution). Table 6 (see Appendix 2) corresponds to problems

SAD08 ($r = 10$, $n = 8$, $m = 12$, $dc = 100$)

saddle value: *58644.621053471*

SAD16 ($r = 10$, $n = 16$, $m = 24$, $dc = 100$)

saddle value: *31142.996423246*

SAD32 ($r = 10$, $n = 32$, $m = 48$, $dc = 100$)

saddle value: *-1200372.0857410*

The control parameter λ of the method was set to 0.5; the process was terminated when the current gap $\Delta(i)$ was reduced to 10^{-6} (in relative accuracy).

Note that theoretically $f(x_i,y_i)$ should not converge to the value of the game (recall that all we claim is that $\varepsilon(x_i^*,y_i^*)$ tends to 0 at the rate prescribed by the theoretical efficiency estimate). Nevertheless, our tests demonstrate that the values $f(x_i,y_i)$ also behave themselves well.

6.3. Constrained minimization

We ran both methods of Sect. 4, i.e., CLM and CNM, on two sets of test problems. Problems of the first set were randomly generated problems of the form

minimize

$$f(x) = (c,x)$$

subject to

$$f_i(x) = \|Q_i x - q_i\|_2 - \rho_i \le 0,\ 1 \le i \le m,$$

$$A_1\, x = b_1,\ A_2 x \le b_2,\ \|x\|_\infty \le r,$$

where x is n-dimensional, Q_i are $k{\times}n$ matrices, and A_1, A_2 are $m_e{\times}n$ and $m_i{\times}n$ matrices, respectively.

The random problems of the above type were created by a simple generator; the input to the generator includes the sizes $(n,\ m,\ k,\ m_e,\ m_i)$, as well as r (size of the box) and the additional control parameters m_{ai}, m_{an} (the numbers of linear inequality constraints and nonlinear constraints active at the solution) and c, dc, ag (responsible for the condition numbers of Q_i, for the range of Lagrange multipliers at the solution and for the range of values of the constraints nonactive at the solution, respectively).

Tables 7 and 8 (see Appendix 2) represent the behaviour of CLM and NLM on two instances

RAND20 $(n = 20,\ m = 8,\ m_e = 2,\ m_i = 4,\ m_{ai} = 2,\ m_{an} = 4,\ k = 10,$ $r = 100,\ c = 10,\ dc = 10,\ ag = 0.1)$

optimal value: *515.95506279904*

RAND40 $(n = 40,\ m = 16,\ m_e = 4,\ m_i = 8,\ m_{ai} = 4,\ m_{an} = 8,\ k = 20,$ $r = 100,\ c = 10,\ dc = 10,\ ag = 0.01)$

optimal value: *-5094.6311010407*

Test problems of the second type were as follows. Consider a chain made of n weightless segments in the vertical plane, and assume that the first segment starts at $(0,0)$ and the last ends at the point $(L,0)$ (the x-axis is horisontal, the y-axis is vertical). The length of each segment is $l = c|x|/n$. At the end of the i-th segment (or, which is the same, at the beginning of the the $(i+1)$-th segment) there is a unit mass, and we minimize the potential energy of the resulting system. In other words, we should minimize the function

$$\sum_{i=1}^{n-1} y_i$$

under the constraints

$$(x_i-x_{i+1})^2 + (y_i-y_{i+1})^2 \le l^2,\ 0 \le i \le n-1,$$

where $x_0 = y_0 = y_n = 0,\ x_n = L.$

The above problem is defined by the data n, L, c. The results in Table 9 (see Appendix 2) correspond to the problems **CHAIN20** ($n = 20,\ c = 2,\ L = 1$) and **CHAIN40** ($n = 40,\ c = 2,\ L = 2$).

[Au. 1972] A. Auslender: Problèmes de Minimax via l'Analyse Convexe et les Inégalités Variationnelles: Théorie et Algorithmes. *Lecture Notes in Economics and Mathematical Systems,* Springer (1972)

[CG 1959] E.W. Cheney, A.A. Goldstein: Newton's methods for convex programming and Tchebytcheff approximation. Numerische Mathenatik 1 (1959) 253-268

[Er. 1966] Yu.M. Ermoliev: Methods for solving nonlinear extremal problems. Cybernetics No. 4 (1966) 1-17

[HK 1971] M. Held, R.M. Karp: The traveling-salesman problem and minimum spanning trees: Part II. Mathematical Programming 1 (1971) 6-25.

[Ke. 1960] J.E. Kelley: The cutting plane method for solving convex programs. Journal of the SIAM 8 (1960) 703-712.

[Ki. 1983] K.C. Kiwiel: An aggregate subgradient method for nonsmooth convex minimization. Mathematical Programming 27/3 (1983) 320-341.

[Ki. 1990] K.C. Kiwiel: Proximity control in bundle methods for convex nondifferentiable minimization. Mathematical Programming 46,1 (1990) 105-122.

[Lb. 1977] B. Yu. Lebedev: On the convergence of the method of loaded functional as applied to a convex programming problem. Journal of Numerical Mathematics and Math. Physics 12 (1977) 765-768 (in Russian).

[Le. 1975] C. Lemaréchal: An extension of Davidon methods to non-differentiable problems. Mathematical Programming Study 3 (1975) 95-109.

[Le. 1978] C. Lemaréchal: Nonsmooth optimization and descent methods. Research Report 78-4 (1978) IIASA, 2361 Laxenburg, Austria.

[LSB 1981] C. Lemaréchal, J.J. Strodiot, A. Bihain: On a bundle algorithm for nonsmooth optimization. In: *Nonlinear Programming 4*, O.L. Mangasarian, R.R. Meyer, S.M. Robinson (eds.), Academic Press (1981) 245-282

[MD 1989] P. Marcotte, J.P. Dussault: A sequential linear programming algorithm for solving monotone variational inequalities. SIAM Journal on Control and Optimization 27,6 (1989) 1260-1278.

[Mi. 1982] R. Mifflin: A modification and an extension of Lemarechal's algorithm for nonsmooth minimization. Mathematical Programming Study 17 (1982) 77-90.

[NYu. 1983] A.S. Nemirovskij, D.B. Yudin: *Problem compexity and method efficiency in optimization*. Wiley-Interscience (1983).

[PD 1978] B.N. Pshenichny, Yu.M. Danilin: *Numerical methods for extremal problems*. Moscow: Mir (1978)

[Pow. 1983] ZQPCVX, a Fortran subroutine for convex programming. Report NA17 (1983), DAMTP, Silver Street, Cambridge CB3 9EW, England.

[Ro. 1972] S.M. Robinson: Extension of Newton's method to nonlinear functions with values in a cone. Numerische Mathematik 9 (1972) 341-347.

[Rock. 1970] R.T. Rockafellar: On the maximality of sums of nonlinear monotone operators. Trans. Amer. Math. Soc. 149 (1970) 75-88.

[SZ 1991] H. Schramm, J. Zowe: A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results. SIAM Journal on Optimization (to appear).

[Wl. 1975] P. Wolfe: A method of conjugate subgradients for minimizing nondifferentiable functions. Mathematical Programming Study 3 (1975) 145-173.

Appendix 1

Let Q be a closed convex subset in $\mathbb{R}^n$ with a nonempty interior, and let F be a monotone mapping with the domain Dom$\{F\}$, int $Q \subseteq$ Dom$\{F\} \subseteq Q$. We establish relations between the following two notions: a solution to the *variational inequality* associated with (F,Q) is a point $x^* \in Q \cap$ Dom$\{F\}$ satisfying

$$(\xi, x-x^*) \geq 0 \text{ for some } \xi \in F(x^*) \text{ and all } x \in Q. \tag{A.1}$$

We define a *weak solution* of the same variational inequality as a point $x^* \in Q$ such that

$$(\eta, x-x^*) \geq 0 \text{ for all } x \in Q \cap \text{Dom}\{F\} \text{ and all } \eta \in F(x). \tag{A.2}$$

Theorem. Let F and Q be defined as above.

Every solution to the variational inequality associated with (F,Q) is a weak solution to this inequality.

Conversely, assume that either

(i) Dom$\{F\} \supseteq Q$ and F is single-valued continuous on Q,

or

(ii) F is maximal monotone.

Then every weak solution to the variational inequality associated with (F,Q) is a solution to this inequality.

Proof. Let $x^* \in Q$ and $\xi \in F(x^*)$ satisfy (A.1). From the monotonicity of F, we have for all $x \in Q \cap$Dom$\{F\}$ and all $\eta \in F(x)$

$$(\eta, x-x^*) \geq (\xi, x-x^*) \geq 0.$$

Let now x^* satisfy (A.2).

For every $y \in Q$ we have $\langle F(x^*+t(y-x^*)), y-x^*\rangle \geq 0$, $0 < t \leq 1$, so that in the case of (i) the continuity of F implies $\langle F(x^*), y-x^*\rangle \geq 0$, $y \in Q$, so that x^* is a solution to the inequality defined by (F,Q).

Now assume that F is maximal monotone on its domain. Consider the normal monotone operator $N(x)$, Dom$\{N\}$ = Q, defined as

$$N(x) = \{\eta \mid \langle\eta,x-y\rangle \geq 0,\ y \in Q\},\ x \in Q.$$

It is well-known that this operator is maximal monotone (recall that Q is a closed convex domain). Now consider the sum $S = N+F$ (Dom$\{S\}$ = Dom$\{F\}\cap$Dom$\{N\}$, $S(x) = \{\eta+\xi \mid \eta \in N(x),\ \xi \in F(x)\}$, $x \in$ Dom$\{S\}$). Since both F and N are maximal monotone and the interiors of their domains have a nonempty intersection int Q, S is maximal monotone (see [Rock. 1970]). If $y \in$ Dom$\{S\}$ and $\zeta \in S(y)$, then $\zeta = \eta + \xi$ for certain $\eta \in N(y)$, $\xi \in F(y)$. We have $\langle\eta,y-x^*\rangle \geq 0$ (since $x^* \in Q$ and in view of the definition of N) and $\langle\xi,y-x^*\rangle \geq 0$ (since x^* is a weak solution to the inequality defined by (F,Q)). It follows that $\langle\zeta,y-x^*\rangle \geq 0$. Thus, x^* is a weak solution to the inequality defined by (S,Q). This fact, in view of Dom$\{S\} \subseteq Q$, means precisely that adding the pair $(x^*,0)$ to the graph of S preserves the monotonicity, and since S is maximal monotone, we conclude that $(x^*,0)$ belongs to the graph of S. Thus, $x^* \in$ Dom$\{F\}$ and there exists $\xi \in F(x^*)$ such that $-\xi \in N(x^*)$. The latter relation means that $\langle\xi,x-x^*\rangle = \langle-\xi,,x^*-x\rangle \geq 0$, $x \in Q$, so that x^* is a solution to the inequality defined by (F,Q). ■

BADGUY30
$f_*(0) = -5000$

#f/g	function
1	-1792.
32	-1867.
33	-1941.33
62	-2015.999
63	-2034.666
64	-2034.666
65	-2039.333
94	-2044.583
95	-2045.312
96	-2045.494
97	-2045.540
98	-2045.551
99	-2045.554
100	-2045.555
101	-2045.555
102	-2045.555
103	-2045.555
104	-2045.555
105	-2045.555
106	-2045.555
107	-2045.555
108	-2045.555
109	-2045.555
110	-2045.555
112	-2045.555
120	-2045.555
124	-2047.111
125	-2047.694
126	-2047.840
129	-2047.876
157	-2047.948
159	-2047.958
160	-2047.961
161	-2047.961
162	-2047.961
163	-2047.961
164	-2047.961
165	-2047.961
166	-2047.961
167	-2047.961
168	-2047.961
169	-2047.961
170	-2047.961
171	-2047.961
172	-2047.961
179	-2047.961
187	-2047.986
188	-2047.995
189	-2047.997
192	-2047.998
220	-2047.999

MAXQUAD
$f_*(0) = -10$

#f/g	function
1	5337.066
2	2663.905
3	1327.046
4	658.5464
5	324.2790
6	157.1409
7	98.60842
8	51.92933
9	28.18153
11	18.12639
12	8.950693
13	4.668303
15	2.387000
18	0.7462724
32	0.5202543
33	-0.5763271
43	-0.6935995
48	-0.7259131
49	-0.7712059
50	-0.8151109
55	-0.8164922
56	-0.8249957
57	-0.8365571
59	-0.8382780
62	-0.8397590
63	-0.8408527
64	-0.8409604
73	-0.8411514
74	-0.8411876
75	-0.8413011
77	-0.8413429
78	-0.8413639
79	-0.8413671
80	-0.8413694
81	-0.8413918
87	-0.8413928
88	-0.8414003
89	-0.8414029
90	-0.8414030
92	-0.8414064
95	-0.8414064
97	-0.8414069
98	-0.8414077

MAXANAL ($\varepsilon=10^{-3}$)
$f_*(0) = -10$

#f/g	function
1	5337.035
2	2663.891
3	1327.039
4	658.5440
5	324.2789
6	157.1426
7	98.59762
8	51.92793
9	28.90709
11	18.34281
12	8.963797
13	4.716963
15	2.333794
17	0.8354944
30	0.6648366
31	0.6388888
33	-0.0013159
34	-0.3767172
40	-0.5076301
42	-0.5510089
43	-0.6109729
44	-0.7338624
51	-0.7360472
52	-0.7887634
53	-0.7961707
54	-0.8100751
56	-0.8103293
57	-0.8225036
58	-0.8289160
62	-0.8299909
64	-0.8304996
66	-0.8306314
75	-0.8307534
87	-0.8307792
91	-0.8307945
95	-0.8307994
97	-0.8308066
102	-0.8308067
103	-0.8308072
104	-0.8308082
108	-0.8308082
110	-0.8308084

TR48
$f_*(0) = -700000.$

#f/g	function
1	-464816.
2	-495706.
3	-520884.
4	-541830.
5	-560801.
6	-562650.
7	-563643.
8	-568830.
9	-578219.
11	-589969.
14	-591689.
17	-598044.
19	-598607.
20	-602712.
22	-603220.
25	-607083.
26	-609600.
28	-613822.
31	-620021.
34	-622699.
35	-626303.
39	-627921.
42	-629003.
43	-630209.
45	-630926.
46	-632947.
48	-633212.
51	-633522.
53	-634393.
56	-634959.
57	-635256.
60	-636015.
61	-636537.
67	-637023.
71	-637073.
72	-637373.
77	-637415.
78	-637520.
80	-637785.
86	-637886.
87	-637886.
88	-637978.
89	-638075.2
91	-638097.9
92	-638148.8
93	-638178.1
94	-638259.0
97	-638283.7
98	-638334.6
101	-638343.7
102	-638392.0
104	-638397.8
105	-638423.4
108	-638468.6
116	-638472.6
117	-638484.8
119	-638486.9
120	-638500.1
124	-638506.8
125	-638531.6
126	-638548.3
127	-638556.6
128	-638560.8
129	-638562.9
130	-638564.0

NET22h
$(10^{-6} \leqslant x)$
$f_*(0) = -200.$

#f/g	function
1	1121.34
2	520.610
3	250.115
4	180.318
5	72.442
7	52.96013
9	4.51785
11	-5.90046
12	-46.08817
16	-61.16107
28	-77.76272
32	-78.53425
33	-83.67625
34	-85.48833
35	-94.05240
36	-95.05838
39	-95.23579
40	-98.86736
41	-100.65651
55	-101.54585
59	-102.06443
63	-102.50231
64	-102.80661
82	-102.89126
96	-102.94307
99	-102.96474
100	-103.11303
108	-103.18887
113	-103.25048
121	-103.25326
122	-103.30134
126	-103.34170
141	-103.34511
143	-103.35174
144	-103.35659
146	-103.37712
149	-103.38933
152	-103.38986
153	-103.39267
155	-103.39548
156	-103.39908
157	-103.40671
197	-103.40673
198	-103.40676
199	-103.40935
204	-103.40938
207	-103.40975
208	-103.41010
209	-103.41055
245	-103.41068
248	-103.41094
257	-103.41106
262	-103.41127
264	-103.41133
270	-103.41148
281	-103.41151
284	-103.41155
286	-103.41157
288	-103.41173
304	-103.41183
306	-103.41190
311	-103.41192
315	-103.41196
321	-103.41198

URYconv
$(-0.2 \leqslant x \leqslant 0.2)$
$f_*(0) = 0.$

#f/g	function
1	10814.
2	5717.
3	3122.
4	1886.6
5	1811.0
8	1412.34
9	1351.46
10	1341.86
11	1255.478
16	1242.176
17	1231.154
19	1227.923
20	1222.121
21	1221.392
22	1218.168
26	1215.048
34	1215.034
35	1214.462
36	1214.244
39	1213.287
45	1213.034
46	1212.893
47	1211.918
50	1211.724
52	1211.495
55	1211.079
58	1210.598
69	1210.587
70	1210.400
77	1210.364
79	1210.343
80	1210.336
82	1210.231
98	1210.169
100	1210.149
105	1210.120
107	1210.100
114	1210.095
118	1210.020
167	1210.019
170	1210.018
171	1210.001
179	1209.998
183	1209.995
207	1209.984
210	1209.963
212	1209.927
218	1209.923
226	1209.915
271	1209.914
297	1209.907
323	1209.903
325	1209.902
342	1209.899
349	1209.896

max. iter = 350

URYconv
(box penalized)
$f_*(0) = 0.$

#f/g	function
1	10814.
2	5717.
3	3122.
4	1886.6
5	1811.0
9	1567.8
56	1519.8
58	1403.6
59	1386.8
60	1306.7
63	1277.5
101	1275.2
107	1274.8
123	1272.5
124	1269.8
166	1267.8
167	1264.9
168	1257.8
170	1255.9
171	1254.0
173	1252.4
175	1241.2
188	1228.8
196	1223.5
199	1221.7
204	1218.1
213	1218.1
221	1216.8
225	1215.9
227	1215.1
298	1215.1
299	1214.7
302	1214.4
310	1213.7
312	1213.5
314	1212.7
321	1212.5
331	1212.5
336	1212.4
339	1212.3
341	1212.1
364	1212.1
383	1212.0
386	1211.7
406	1211.7
422	1211.7
430	1211.7
431	1211.6
432	1211.6
440	1211.5
447	1211.3
451	1211.2
453	1211.2
465	1211.2
481	1211.2
486	1211.1
488	1211.1
490	1211.0
497	1210.9

max. iter = 500

TSP6

$f_*(0) = -1000$

#f/g	function
1	-403.
2	-416.75
3	-472.00
5	-611.50
9	-612.9643
10	-614.5168
11	-616.2584
12	-617.0000

TSP14

$f_*(1) = -4000.$

#f/g	function
1	-2633.
2	-2721.
3	-2934.729
4	-3181.616
10	-3187.119
11	-3200.685
12	-3226.135
13	-3259.120
15	-3301.031
19	-3313.501
22	-3317.878
23	-3320.689
24	-3321.485
25	-3322.000

TSP29

$f_*(0) = -3000.$

#f/g	function
1	-1666.
16	-1756.8
17	-1765.9
18	-1769.9
19	-1877.0
21	-1880.5
22	-1932.03
25	-1963.80
30	-1965.26
31	-1984.24
35	-1996.98
41	-1998.32
43	-2002.882
47	-2004.106
49	-2005.646
51	-2006.982
52	-2010.877
53	-2012.807
54	-2013.013
55	-2013.080
56	-2013.151
61	-2013.199
62	-2013.329
63	-2013.415
64	-2013.457
65	-2013.478
66	-2013.489
67	-2013.495
68	-2013.497

TSP100

$f_*(0) = -30000.$

#f/g	function
1	-18993.07
10	-19161.91
11	-19858.97
15	-19954.67
18	-20488.03
33	-20568.25
42	-20598.64
43	-20710.57
51	-20749.55
61	-20784.17
65	-20792.46
67	-20809.70
72	-20873.96
84	-20882.46
99	-20882.69
101	-20898.59
104	-20910.15
106	-20914.27
107	-20914.60
108	-20922.23
127	-20923.17
128	-20925.51
129	-20928.94
134	-20929.85
135	-20931.04
136	-20932.95
137	-20933.32
138	-20935.30
139	-20935.75
141	-20936.03
142	-20936.06
143	-20936.49
144	-20936.76
145	-20937.06
146	-20937.22
147	-20937.48
148	-20937.63
149	-20937.73
150	-20937.81
151	-20937.87
152	-20937.91

TSP120

$f_*(0) = -8000.$

#f/g	function
1	-5840.
2	-6074.048
3	-6240.566
4	-6308.962
6	-6403.346
7	-6481.775
30	-6578.004
31	-6587.233
33	-6633.019
36	-6647.750
38	-6678.937
44	-6694.561
47	-6737.301
51	-6757.920
55	-6775.514
72	-6779.020
75	-6794.310
79	-6799.058
80	-6803.682
82	-6812.904
95	-6841.046
102	-6858.842
107	-6858.956
112	-6866.910
125	-6874.902
133	-6878.229
136	-6881.725
140	-6887.902
150	-6892.160
160	-6893.193
162	-6894.098
169	-6896.184
174	-6897.147
175	-6898.829
177	-6900.010
184	-6900.310
186	-6901.119
189	-6902.833
196	-6905.196
206	-6905.214
207	-6906.310
212	-6906.700
214	-6906.944
215	-6907.558
218	-6907.634
219	-6908.053
221	-6908.970
231	-6909.201
235	-6909.221
237	-6909.729
240	-6909.925
244	-6910.158
245	-6910.160
246	-6910.327
253	-6910.494
254	-6910.509
255	-6910.773
260	-6910.864
261	-6910.988
271	-6911.009
272	-6911.019
273	-6911.096
274	-6911.113
275	-6911.132
276	-6911.150
277	-6911.172
278	-6911.190
279	-6911.199
281	-6911.211
282	-6911.219
283	-6911.225
284	-6911.232
285	-6911.234
286	-6911.238
287	-6911.241
288	-6911.246

TSP442

$f_*(0) = -60000.$

#f/g	function
1	-46862.30
19	-47083.30
21	-47754.42
23	-48064.40
27	-48314.50
31	-48452.01
33	-48464.14
35	-48545.13
37	-48584.86
38	-48740.62
40	-48763.24
41	-49131.83
46	-49154.38
48	-49176.18
51	-49230.08
56	-49334.63
59	-49412.26
62	-49416.44
63	-49513.38
67	-49674.64
72	-49745.21
76	-49773.25
79	-49815.93
85	-49827.59
87	-49883.24
92	-49910.34
95	-49917.70
104	-50034.19
112	-50045.58
116	-50099.89
121	-50110.86
123	-50142.61
139	-50149.73
141	-50164.50
145	-50182.23
150	-50198.06
154	-50206.96
160	-50215.61
163	-50235.17
167	-50236.76
168	-50259.73
173	-50263.81
174	-50292.01
179	-50297.04
182	-50312.75
186	-50326.23
192	-50332.78
196	-50335.65
210	-50345.95
219	-50349.48
222	-50374.53
235	-50384.63
257	-50386.45
258	-50390.79
262	-50400.22
269	-50408.08
279	-50415.16
283	-50422.02
286	-50427.40
296	-50437.49
319	-50437.75
329	-50438.35
349	-50444.87
353	-50462.73
384	-50466.24
389	-50467.07
393	-50471.39
396	-50471.77
401	-50474.84
405	-50475.85
410	-50477.97
412	-50480.37

max. iter = 420

SAD08

#f/g	gap	objective
2	208364.1	26892.39
3	89255.71	46355.59
4	43520.98	52638.27
5	18244.58	55997.28
6	7741.616	57240.14
7	3532.174	58045.98
8	1523.933	58236.56
9	957.308	58501.01
10	470.576	58525.92
11	277.448	58604.07
12	158.842	58579.83
13	113.201	58627.24
14	40.343	58634.34
15	30.218	58637.45
16	19.760	58642.40
17	5.872	58643.52
18	3.835	58644.10
19	2.722	58643.97
20	1.754	58644.54
21	0.751	58657.39
22	0.421	58651.07
23	0.313	58647.68
24	0.289	58646.00
25	0.282	58645.22
26	0.271	58644.77
27	0.268	58644.64
28	0.264	58644.71
29	0.259	58644.65
30	0.249	58644.68
31	0.240	58644.66
32	0.230	58644.66
33	0.219	58644.64
34	0.205	58644.62
35	0.167	58644.61
36	0.083	58644.60

SAD16

#f/g	gap	objective
2	117494.	34661.43
3	18235.	33525.44
4	9828.8	31997.96
5	4291.3	31609.38
6	2361.1	31285.99
7	1388.1	31522.31
8	1066.3	31259.49
9	692.5	31236.11
10	454.9	31521.22
11	395.7	31231.72
12	275.7	31177.37
13	203.1	31181.35
14	163.0	31190.12
15	109.4	31159.29
16	76.94	31178.15
17	63.31	31150.13
18	44.27	31152.80
19	35.00	31148.70
20	24.12	31144.33
21	20.91	31145.55
22	16.96	31144.76
23	10.62	31142.88
24	7.97	31143.32
25	6.34	31143.28
26	4.69	31144.14
27	4.47	31143.33
28	3.91	31143.80
29	3.01	31143.05
30	2.65	31143.20
31	2.18	31143.22
32	1.66	31143.23
33	1.41	31143.10
34	0.92	31143.37
35	0.72	31142.97
36	0.47	31143.05
37	0.36	31143.04
38	0.20	31142.92
39	0.12	31142.99
40	0.10	31142.99
41	0.08	31142.99
42	0.06	31143.01
43	0.05	31143.00

SAD32

#f/g	gap	cost
2	5448581.	-963480.
3	1121204.	-1202545.
4	458570.	-1334416.
5	362633.	-1218794.
6	205589.	-1207426.
7	138308.	-1208932.
8	99714.	-1202523.
9	74153.	-1204505.
10	30586.	-1204660.
11	27454.	-1202606.
12	22629.	-1202367.
13	16673.	-1208258.
14	13224.	-1200468.
15	11392.	-1201155.
16	8064.8	-1200846.
17	5408.5	-1203882.
18	5062.2	-1199727.
19	3861.4	-1200660.
20	3216.3	-1200496.
21	2510.4	-1200619.
22	2064.3	-1200269.
23	1584.9	-1200410.
24	1395.2	-1200444.
25	907.70	-1200245.
26	699.11	-1200454.
27	640.95	-1200440.
28	552.07	-1200428.
29	441.66	-1200394.
30	342.32	-1200384.
31	269.02	-1200385.
32	194.68	-1200375.
33	180.54	-1200384.
34	136.27	-1200415.
35	120.27	-1200366.
36	88.83	-1200375.
37	72.57	-1200391.
38	70.28	-1200368.
39	66.52	-1200372.
40	39.70	-1200375.
41	30.54	-1200382.
42	28.96	-1200373.
43	23.00	-1200374.
44	23.81	-1200375.
45	16.84	-1200373.
46	10.23	-1200374.
47	9.07	-1200373.
48	8.99	-1200373.
49	7.21	-1200373.
50	5.23	-1200372.
51	4.93	-1200375.
52	3.26	-1200372.
53	3.27	-1200372.
54	2.51	-1200372.
55	2.11	-1200372.
56	2.78	-1200372.
57	1.94	-1200372.
58	1.02	-1200372.
59	1.65	-1200371.
60	1.51	-1200372.

RAN20 (constrained level)

$f_*(0) = 0.$

#f/g	objective	infeasibility
2	-3104.446	1245.2
3	-5627.954	1035.0
4	-7415.139	833.
5	-7890.228	721.0
5	-7890.228	721.0
6	-2975.067	605.21
7	-902.1215	154.4
8	593.0856	48.44
9	595.1046	16.85
10	573.9980	9.348
11	562.2839	2.424
12	562.1910	0.7538
13	565.0481	0.3537
14	558.4729	-0.0331
15	557.0874	-0.0529
16	544.9237	-0.1035
17	543.6074	-0.1761
18	533.5708	-0.2709
19	521.0267	0.01384
20	521.2570	-0.00166
21	521.3004	-0.00457
22	519.5250	-0.0859
23	516.9066	-0.00577
24	516.6392	-0.01011
25	516.5071	-0.003914
26	516.3140	-0.003918
27	516.2809	-0.003636
28	516.2903	-0.003980
29	516.0568	-0.000514
30	516.0311	-0.000610
31	516.0270	-0.000408
32	516.0190	-0.001225
33	516.0146	-0.001059
34	515.9704	-0.000190
35	515.9644	-0.000023
36	515.9634	-0.000036
37	515.9583	-0.000063
38	515.9582	-0.000040
39	515.9569	-0.000042
40	515.9559	-0.000015
42	515.9558	-0.000012
43	515.9558	-0.000013
44	515.9558	-0.000013
45	515.9558	-0.000013
46	515.9557	-0.000008
47	515.9554	-0.000007

RAN20 (Newton level)

$f_*(0) = 0.$

#f/g	objective	infeasibility
2	44.59312	326.96723
3	32.69669	146.02230
4	14.60221	68.737178
6	21.58758	40.448920
7	26.23598	38.418595
8	29.03952	34.057029
9	29.71941	31.842534
10	419.1160	7.2074009
11	475.3577	3.9754218
12	501.9541	1.3725472
13	514.9950	0.5491130
15	514.4676	0.2290687
18	513.1268	0.2275991
19	514.0681	0.1404776
21	515.6814	0.0352848
22	516.0958	0.0092386
26	515.8369	0.0053595
28	516.1400	0.0031480
29	516.0326	0.0026109
30	516.0124	0.0015542
32	515.9719	0.0003822
37	515.9474	0.0003219
39	515.9553	0.0000367
40	515.9549	0.0000249
41	515.9546	0.0000237
42	515.9549	0.0000085
47	515.9551	0.0000061
48	515.9550	0.0000047
49	515.9550	0.0000056
50	515.9551	0.0000012
53	515.9551	0.0000003

RAN40 (constrained level)
$f_*(0) = -10000.$

#f/g	objective	infeasibility
2	-6183.871	782.028
3	-5187.675	193.685
4	-4938.308	49.4309
5	-5060.899	12.6758
6	-5084.405	5.27870
7	-5037.375	2.51922
8	-5069.418	1.12415
9	-5053.608	0.51602
10	-5068.662	0.13841
11	-5069.223	0.07101
12	-5067.736	0.02210
13	-5067.736	0.02215
14	-5067.736	0.02216
15	-5067.733	0.02241
16	-5067.732	0.02251
17	-5078.722	-0.01970
18	-5079.693	-0.00561
19	-5079.723	-0.00385
20	-5082.975	-0.15849
21	-5087.995	-0.06667
22	-5090.095	-0.02230
23	-5090.089	-0.02249
24	-5093.158	-0.00964
25	-5093.157	-0.00967
26	-5094.048	-0.00276
27	-5094.203	-0.00168
28	-5094.202	-0.00170
29	-5094.202	-0.00171
30	-5094.322	-0.00275
31	-5094.349	-0.00284
32	-5094.570	-0.00078
33	-5094.570	-0.00078
34	-5094.570	-0.00078
35	-5094.570	-0.00078
36	-5094.567	-0.00056
37	-5094.624	-0.00002
38	-5094.625	-0.00002
39	-5094.625	-0.00002
40	-5094.624	-0.00002
41	-5094.624	-0.00002
42	-5094.625	-0.00002
43	-5094.625	-0.00006

RAN40 (Newton level)
$f_*(0) = -10000.$

#f/g	objective	infeasibility
2	-6134.477	780.30
3	-5167.408	193.91
4	-4924.203	49.429
5	-5062.152	12.665
6	-5081.076	5.3618
7	-5037.342	2.6256
12	-5037.311	1.1368
13	-5044.745	0.6101096
15	-5090.168	0.3883399
16	-5082.450	0.0751529
20	-5087.145	0.0377584
23	-5093.960	0.0120858
24	-5093.625	0.0022751
31	-5094.558	0.0002471
34	-5094.627	0.0000513
40	-5094.621	0.0000099
47	-5094.631	0.0000075
50	-5094.628	0.0000011
54	-5094.631	0.0000004

CHAIN20 (constrained level)

$f_*(0) = -1000.$

#f/g	objective	infeasibility
2	-14.56595	0.6682579
3	-14.07672	0.6425655
4	-13.51848	0.6132536
5	-12.87745	0.5796025
6	-14.54907	0.4726217
7	-19.70359	0.1872318
8	-16.86018	0.1475199
9	-13.83541	0.1055811
10	-10.60577	0.0614231
11	-8.89049	0.0384369
12	-8.98122	0.0448675
13	-9.21371	0.0242676
14	-9.13262	0.0235769
15	-9.17770	0.0074056
16	-9.07817	0.0064999
17	-9.08397	0.0034704
18	-9.07923	0.0033274
19	-9.10023	0.0011768
20	-9.09121	0.0011246
21	-9.10110	0.0005591
22	-9.09984	0.0004791
23	-9.10172	0.0002388
24	-9.10342	0.0000675
25	-9.10398	0.0000117
26	-9.10405	0.0000318
27	-9.10404	0.0000038
28	-9.10408	0.0000016
29	-9.10402	0.0000011
30	-9.10398	0.0000005
31	-9.10398	0.0000006

CHAIN20 (Newton level)

$f_*(0) = -1000.$

#f/g	objective	infeasibility
2	-14.98764	0.6904065
4	-15.96882	0.6783116
5	-18.38541	0.6413603
6	-22.98693	0.5690655
7	-31.24539	0.4371859
8	-38.02470	0.3270592
9	-17.77501	0.1042374
14	-18.50436	0.0921989
15	-9.244903	0.1659744
17	-9.380816	0.0058774
19	-9.414706	0.0031686
20	-9.129002	0.0012095
22	-9.140639	0.0004056
24	-9.103276	0.0001196
25	-9.104479	0.0000123
27	-9.104542	0.0000065
29	-9.103956	0.0000028
30	-9.103983	0.0000007

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique

ISSN 0249-6399